МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВАСИЛЯ СТУСА

КОМАР ОЛЕКСАНДРА ОЛЕГІВНА

Допускається до захисту:
в.о. завідувача кафедри
інформаційних технологій
канд. техн. наук, доцент
_____ О. В. Зелінська
«____»_____ 20__ р.

МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ СПОЖИВАННЯ ГАЗУ КОМУНАЛЬНИМ ПІДПРИЄМСТВОМ НА ОСНОВІ ЧАСОВИХ РЯДІВ

Спеціальність 122 Комп’ютерні науки

Кваліфікаційна (бакалаврська) робота

Керівник:
Т. В. Січко, доцент кафедри
інформаційних технологій,
к. т. н., доцент
_____

Оцінка: _____ / _____ / _____
(бали/за шкалою ЄКТС/за національною шкалою)
Голова ЕК: _____

Вінниця 2025

АНОТАЦІЯ

Комар О.О. Моделювання та прогнозування споживання газу комунальним підприємством на основі часових рядів. Спеціальність 122 «Комп'ютерні науки», освітня програма «Комп'ютерні науки». Донецький національний університет імені Василя Стуса, Вінниця 2025.

У кваліфікаційній (бакалаврській) роботі досліджено процес побудови моделей прогнозування споживання природного газу на основі часових рядів. Проведено порівняння класичних статистичних підходів і сучасних інтелектуальних методів прогнозування. З використанням реальних даних реалізовано моделі Holt-Winters, SARIMA, ANFIS та N-BEATS. За результатами оцінки точності побудовано рекомендації щодо вибору ефективного методу прогнозування.

Ключові слова: прогнозування, часовий ряд, споживання газу, Holt-Winters, SARIMA, ANFIS, N-BEATS.

69 с., 26 рис., 6 табл., 8 дод., 42 джерел.

ABSTRACT

Komar O.O Modeling and forecasting gas consumption by a municipal enterprise based on time series. Specialty 122 «Computer Science», educational program «Computer Science». Vasyl Stus Donetsk National University, Vinnytsia 2025.

In the qualification (bachelor's) work investigates the process of building gas consumption forecasting models based on time series. A comparison of classical statistical methods and modern intelligent forecasting approaches is conducted. Using real data, models such as Holt-Winters, SARIMA, ANFIS, and N-BEATS were implemented. Based on the accuracy assessment results, recommendations for choosing the most effective forecasting method are provided.

Keywords: forecasting, time series, gas consumption, Holt-Winters, SARIMA, ANFIS, N-BEATS.

69 p., 26 fig., 6 tables, 8 add., 42 references.

ЗМІСТ

ВСТУП

В умовах зростаючої нестабільності енергетичного ринку та посилення вимог до економічної ефективності, прогнозування споживання природного газу стає критично важливим елементом діяльності комунальних підприємств. Невизначеність у майбутніх обсягах газоспоживання може призводити до економічних втрат, надлишкових закупівель або нестачі ресурсу, що негативно позначається на фінансовій стабільності підприємства та якості обслуговування споживачів.

У сучасних умовах, коли енергетична безпека та раціональне використання ресурсів в Україні набувають стратегічного значення, зростає потреба в удосконаленні методів прогнозування. Розвиток статистичних, машинних і нейромережевих підходів відкриває нові можливості для підвищення точності прогнозів, що, у свою чергу, дозволяє підприємствам оптимізувати управлінські рішення, ефективніше розподіляти ресурси та мінімізувати витрати.

Вибір оптимального методу прогнозування заслуговує особливої уваги, оскільки від цього напряму залежить точність отриманих результатів та ефективність прийнятих на їх основі рішень. Різні методи моделювання часових рядів мають свої переваги та обмеження в контексті специфіки газоспоживання, що вимагає комплексного підходу до їх відбору, налаштування та оцінки ефективності.

Метою роботи є застосування сучасних методів прогнозування часових рядів для оцінки майбутнього споживання газу комунальним підприємством, а також порівняння цих методів для вибору оптимальної моделі.

Для досягнення поставленої мети у роботі вирішуються такі завдання:

- дослідити основні поняття часових рядів та їх характеристики;
- розглянути сучасні методи прогнозування часових рядів;
- проаналізувати специфіку застосування обраних методів для прогнозування газоспоживання;
- реалізувати математичні моделі у середовищі програмування;

- виконати порівняльний аналіз точності прогнозування для різних моделей за допомогою відповідних метрик оцінки;
- надати рекомендації щодо вибору найбільш ефективного методу прогнозування.

Об’єктом дослідження є процес прогнозування споживання природного газу комунальним підприємством.

Предметом дослідження є математичні моделі прогнозування часових рядів та їх застосування для оцінки майбутнього обсягу газоспоживання.

Теоретичне значення дослідження полягає у систематизації підходів до прогнозування часових рядів у сфері енергетики та аналізі можливостей різних моделей для роботи з такими даними.

Практичне значення полягає у розробці та тестуванні прогнозних моделей, які в подальшому можуть бути використані для підвищення точності оцінки майбутнього споживання газу, що сприятиме ефективнішому управлінню ресурсами на рівні комунального підприємства.

РОЗДІЛ 1
ТЕОРЕТИЧНІ ОСНОВИ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ ГАЗОСПОЖИВАННЯ

1.1 Поняття та основні характеристики часових рядів

У сучасному світі, де прийняття ефективних рішень дедалі частіше базується на аналізі даних, особливу роль відіграють часові ряди – один із ключових інструментів для вивчення динамічних процесів. У багатьох сферах діяльності, зокрема в економіці, енергетиці, медицині, фінансах та екології, дані фіксуються протягом певного періоду, що дозволяє відстежувати зміни показників у часі та виявляти приховані закономірності. Застосування часових рядів дає змогу трансформувати неструктуровані дані у формалізовані аналітичні залежності, які можна використати для підвищення ефективності, оптимізації ресурсів або раннього виявлення відхилень у процесах.

Часовий ряд – це впорядкована у часі послідовність значень певної змінної, зафіксованих через однакові проміжки часу. Такими значеннями можуть бути, наприклад, щоденні обсяги споживання газу підприємством, місячні показники температури повітря або річні дані про врожайність сільськогосподарських культур. На відміну від звичайних вибірок даних, часові ряди мають важливу особливість – порядок спостережень має суттєве значення, оскільки кожне наступне значення може залежати від попередніх або бути з ними корельованим.

Ретельне вивчення часових рядів передбачає аналіз їхніх внутрішніх структурних складових. Це дозволяє краще зрозуміти природу даних, а також застосовувати відповідні моделі для прогнозування. До основних компонентів часового ряду зазвичай відносять такі:

- тренд – це довгострокова тенденція зміни рівня ряду, яка відображає загальний напрямок розвитку показника протягом певного періоду часу. Тренд може бути зростаючим, спадним або стабільним і зазвичай пов’язаний зі стійкими факторами, такими як технологічний прогрес, зростання населення чи зміни в законодавстві;

- сезонність – це регулярні, передбачувані коливання, що повторюються з певною періодичністю, як правило, в межах року. Наприклад, попит на природний газ зростає в зимовий період через опалення, що зумовлює характерну сезонну структуру споживання;
- циклічність – це коливання, що відбуваються протягом триваліших періодів (у порівнянні з сезонністю), пов’язані, зокрема, з економічними циклами. Вони не мають чітко визначеної періодичності, однак часто є результатом змін у політичному чи соціально-економічному середовищі;
- іррегулярна (випадкова) компонента – це непередбачувані флуктуації, які виникають внаслідок унікальних подій або випадкових факторів. Ці відхилення важко моделювати, проте їх наявність слід враховувати при побудові моделей для прогнозування.

Часові ряди також класифікують за способом взаємодії між їхніми компонентами. Основними типами такої структури є:

- адитивне представлення, при якому загальне значення часового ряду формується як сума тренду, сезонності, циклічності та випадкових коливань. Такий підхід доцільно застосовувати, коли амплітуда сезонних коливань залишається сталою впродовж усього періоду спостережень;
- мультиплікативне представлення, при якому значення ряду є добутком його окремих компонент. Цей варіант більш доречний, коли сезонні або циклічні коливання зростають разом зі збільшенням тренду або загального рівня ряду.

Виявлення та оцінка зазначених компонентів відбувається за допомогою низки методів. Так, для згладжування ряду і виявлення тренду застосовують метод ковзного середнього, який дозволяє усунути короткотермінові флуктуації та візуалізувати загальну тенденцію змін. Визначити наявність і силу сезонності допомагає автокореляційна функція (ACF), що показує взаємозв’язок значень ряду на різних часових лагах, дозволяючи виявити повторювані закономірності. Для більш точної оцінки сезонної компоненти також використовують сезонну декомпозицію з використанням STL (Seasonal-Trend decomposition using Loess)

або інші алгоритми розділення ряду на складові. Виявлення випадкових коливань, у свою чергу, можливе після усунення тренду та сезонності, коли залишкова частина ряду характеризується непередбачуваністю. Такий підхід до аналізу дозволяє краще підготувати дані до подальшого моделювання та обрати відповідну модель прогнозування.

Класифікують часові ряди також за частотою спостережень. Залежно від інтервалу збору даних виділяють високочастотні ряди (спостереження через секунди, хвилини або години), середньочастотні (щоденні, щотижневі) та низькочастотні (щомісячні, квартальні, річні). Кожен тип має свої особливості моделювання: високочастотні ряди зазвичай містять більше шуму і потребують фільтрації, а низькочастотні – часто відображають сезонні або довгострокові економічні тренди.

Обсяг доступних даних суттєво впливає на ефективність прогнозування часових рядів. У літературі з класичного статистичного аналізу часто наводяться орієнтовні рекомендації щодо мінімальної кількості спостережень – наприклад, не менше 50 для звичайних рядів та 80–100 для сезонних. Водночас потреба в обсязі даних може істотно варіюватися залежно від обраної моделі: складні методи, зокрема нейронні мережі, зазвичай потребують значно більших вибірок для досягнення стабільного результату [21].

Важливою властивістю часових рядів, що впливає на вибір методів аналізу та моделювання, є стаціонарність. Ряд вважається стаціонарним, якщо його основні статистичні характеристики – середнє значення, дисперсія, автокореляція – залишаються незмінними в часі. У стаціонарному ряді відсутні чітко виражені тренди та сезонні коливання. Нестаціонарні ряди, навпаки, характеризуються зміною цих параметрів упродовж періоду спостереження, що ускладнює побудову математичних моделей та знижує точність прогнозування.

Тому перед моделюванням нестаціонарні ряди часто проходять етап трансформації до стаціонарного вигляду. Це може включати вилучення тренду, усунення сезонності, логарифмування, а також застосування диференціювання – обчислення різниць між послідовними значеннями часового ряду.

Диференціювання особливо часто застосовується в рамках ARIMA-моделей і є базовим інструментом для досягнення стаціонарності без втрати суттєвої інформації про структуру даних. Правильна трансформація ряду є необхідною умовою для коректного використання багатьох методів прогнозування, особливо тих, що засновані на припущенні про стаціонарність.

Знання структури та властивостей часового ряду дозволяє обґрунтовано підходити до вибору методів прогнозування. Адже різні методики по-різному реагують на наявність тренду чи сезонності, і лише правильне врахування цих компонент забезпечує надійність прогнозів.

У зв'язку з цим розроблено різноманітні підходи до роботи з часовими рядами – від класичних статистичних до сучасних інтелектуальних моделей. Основні з них будуть розглянуті у підпункті 1.2, тут буде зазначено лише їх коротку класифікацію. Методи для аналізу та прогнозування часових рядів можна умовно поділити на кілька груп:

• класичні статистичні методи, як-от ковзне середнє, експоненціальне згладжування або декомпозиція часового ряду, широко застосовуються для первинного аналізу та побудови простих прогнозних моделей. Вони дозволяють зменшити вплив шуму, виявити основні закономірності в даних - тренд або сезонність, і побудувати короткострокові прогнози. Ці методи добре працюють у випадках, коли ряд має стабільну структуру та не надто складну динаміку.

• моделі ARIMA (AutoRegressive Integrated Moving Average) є потужним інструментом для аналізу нестаціонарних часових рядів. Вони поєднують в собі компоненти авторегресії (AR), інтегрування (I) та ковзного середнього (MA), що дає змогу моделювати як внутрішню інерційність, так і глобальні тенденції у даних. У разі наявності сезонної складової використовується її розширення – SARIMA. Цей клас моделей передбачає детальну попередню обробку даних, зокрема перевірку стаціонарності та диференціювання, але вважається одним з найнадійніших для побудови інтерпретованих прогнозів.

• методи машинного навчання, а саме дерева рішень, штучні нейронні мережі, градієнтний бустинг та інші ансамблеві підходи, дозволяють виявляти

складні нелінійні залежності в даних, які складно помітити класичними методами. Вони добре масштабуються на великі обсяги даних і здатні автоматично виявляти закономірності, що підвищує точність прогнозування. Водночас ці методи вимагають більшого обсягу навчальних даних і ретельного налаштування гіперпараметрів, а також часто мають обмежену інтерпретованість, що слід враховувати в прикладних задачах.

Однією з поширених проблем під час аналізу часових рядів у сфері енергетики є наявність пропущених значень та аномальних спостережень. Вони можуть виникати внаслідок технічних збоїв у системах обліку, перерв у передаванні даних або помилок під час збору інформації. Подібні дефекти впливають на цілісність ряду, а отже – і на точність прогнозних моделей. У типових випадках для обробки пропусків застосовуються методи лінійної або поліноміальної інтерполяції, згладжування, а також алгоритми, засновані на машинному навчанні. Аномальні значення, що не відповідають загальній структурі ряду, зазвичай виявляють за допомогою міжквартильного розмаху, Z-оцінки або кластерного аналізу.

Часові ряди є складною, але надзвичайно важливою формою представлення динамічних процесів, яка дозволяє не лише фіксувати зміни показників у часі, а й робити обґрунтовані прогнози на основі виявлених закономірностей. Розуміння їхньої структури, типів та властивостей є необхідною умовою для побудови ефективних моделей, що застосовуються в багатьох сферах – зокрема в енергетиці, де своєчасне прогнозування споживання ресурсів має як економічне, так і стратегічне значення. Саме тому доцільно розглянути, чому прогнозування споживання природного газу є таким важливим, і яке значення воно має як для окремих підприємств, так і для енергетичної системи країни загалом.

Прогнозування обсягів споживання природного газу є ключовим елементом у системі управління ресурсами комунального підприємства. Від точності таких прогнозів залежать не лише фінансові результати, а й здатність підприємства забезпечувати безперебійне теплопостачання, планувати закупівлі

палива, ефективно керувати інфраструктурою та відповідати на зміни попиту. За умов енергетичної нестабільності та зростання вимог до енергоефективності, потреба в точних прогнозах стає дедалі актуальнішою.

Часові ряди, що відображають динаміку споживання природного газу, дають змогу аналізувати минулі зміни попиту, виявляти тенденції та сезонні коливання, а також будувати прогнози для оптимізації подальшого використання ресурсу. Для комунальних підприємств це означає зниження ризику дефіциту або надлишкових запасів, зменшення витрат, пов'язаних із неплановими закупівлями, та підвищення надійності надання послуг. Крім того, точне прогнозування є важливим інструментом у переговорах із постачальниками газу, дозволяючи укладати контракти на більш вигідних умовах та уникати небажаних фінансових наслідків.

Хоча у роботі моделювання здійснюється на прикладі конкретного комунального підприємства, підхід до прогнозування, заснований на часових рядах, має універсальний характер. За умови якісного налаштування, подібні моделі може бути підлаштовано для потреб інших установ або територіальних енергетичних систем, оскільки спираються на фундаментальні властивості динаміки споживання.

Водночас важливість точного прогнозування виходить далеко за межі інтересів окремих підприємств. Для України загалом це є основою сталого функціонування газотранспортної системи, ефективного планування закупівель енергоносіїв і забезпечення енергетичної безпеки. Відповідно до Стратегії енергетичної безпеки України, серед пріоритетних завдань визначено забезпечення надійності постачання природного газу, диверсифікацію джерел енергії та розвиток прогнозних механізмів для запобігання кризовим ситуаціям у галузі [3].

У межах цього дослідження прогнозування здійснюється виключно на основі часових рядів фактичного споживання природного газу, без використання додаткових змінних, таких як температура повітря, тарифи чи макроекономічні індикатори. Такий підхід обрано з огляду на практичність, доступність

історичних даних і прагнення до універсальності моделі. У центрі уваги – внутрішня динаміка самого ряду: трендові зміни та сезонні коливання, які відіграють ключову роль у формуванні споживчого профілю. Ці компоненти широко використовуються в класичних і сучасних підходах до прогнозування, оскільки дозволяють ефективно моделювати типові повторювані закономірності, такі як сезонний попит чи поступове зростання споживання. Врахування лише внутрішніх характеристик дозволяє створити компактну, гнучку та універсальну модель, яка може бути швидко адаптована до нових даних і застосована в умовах обмеженої інформації, без втрати точності прогнозу.

У підсумку, прогнозування споживання природного газу виступає важливим інструментом як на рівні локального управління ресурсами, так і в контексті державної енергетичної політики. Моделі часових рядів дозволяють виявляти закономірності в динаміці споживання та адаптуватися до змін, що робить їх ефективним інструментом для прийняття обґрунтованих рішень. У наступному підпункті розглянуто сучасні методи прогнозування часових рядів, які застосовуються для реалізації цієї задачі.

1.2 Сучасні методи для прогнозування споживання природного газу

Прогнозування обсягів споживання природного газу є багаторівневим завданням, яке потребує урахування особливостей поведінки даних у часі та вибору відповідного підходу для побудови точних прогнозів. У цьому контексті важливо ознайомитися з методами, які найчастіше застосовуються для моделювання динаміки споживання енергоресурсів, зокрема природного газу. Сучасна практика аналізу даних передбачає використання широкого спектра методів – від класичних статистичних моделей, що добре працюють із простою сезонною динамікою, до складних алгоритмів штучного інтелекту, здатних виявляти приховані нелінійні залежності. Вибір конкретного підходу до прогнозування визначається структурою наявних даних, тривалістю історичних спостережень, цілями аналізу, очікуваною точністю та технічними можливостями реалізації моделі в умовах реального застосування.

Одним із найпоширеніших підходів у практиці прогнозування є метод експоненційного згладжування, зокрема його модифікація у вигляді троїстого згладжування, відома як модель Хольта-Вінтерса. Це класичний статистичний підхід, що дозволяє враховувати тренд і сезонність, що робить його зручним для задач енергетики, де споживання природного газу має яскраво виражену сезонну складову. У своїй базовій формі модель передбачає незалежне оцінювання трьох компонент: рівня, тренду та сезонності, що робить її прозорою у налаштуванні й інтерпретації. Завдяки відносній простоті й низьким обчислювальним витратам, вона часто використовується як базовий інструмент для створення оперативних прогнозів. Проте ефективність моделі значно знижується у випадках, коли структура ряду змінюється з часом або присутні складні взаємозв'язки між значеннями.

У таких ситуаціях доцільно застосовувати інший класичний статистичний підхід – авторегресійні моделі з інтегрованим ковзним середнім, які дають змогу моделювати залежність між спостереженнями на різних часових лагах. Найбільш відомою є модель SARIMA (Seasonal ARIMA), яка розширює класичну ARIMA шляхом включення сезонних параметрів, що надає змогу ефективно моделювати як коротко-, так і довготривалі сезонні коливання. Завдяки гнучкій структурі, яка включає авторегресивні, інтегруючі та ковзні складові, ця модель адаптується до широкого спектра задач прогнозування, особливо тоді, коли часовий ряд демонструє нестаціонарність. Водночас, для ефективного використання SARIMA необхідна якісна попередня обробка даних, включно з перевіркою стаціонарності та підбором параметрів, що може бути досить трудомістким у порівнянні з простішими методами.

Окреме місце серед сучасних підходів займають методи машинного навчання, які дозволяють виявляти приховані закономірності у даних, що не завжди доступні при використанні класичних алгоритмів. Вони базуються на здатності моделі самостійно навчатися з даних, знаходячи складні нелінійні взаємозв'язки, які важко формалізувати вручну. Наприклад, методи на основі дерев рішень, ансамблеві моделі, такі як градієнтний бустинг, або ж прості

нейронні мережі часто забезпечують високу точність прогнозів за умови наявності великої кількості якісних вхідних даних. Завдяки своїй адаптивності та можливості реагувати на зміни у структурах даних, машинне навчання все частіше використовується у сфері енергетичного прогнозування. Проте ці методи мають певні вимоги до обсягу даних і можуть бути чутливими до перевищення складності моделі, що вимагає додаткової уваги до вибору гіперпараметрів та регуляризації.

Окрім традиційних математичних підходів, сучасні дослідження звертають увагу на використання нечіткої логіки як альтернативного підходу до моделювання складних систем. Поєднання систем нечітких висловлювань із методами машинного навчання надає можливість створювати адаптивні алгоритми, здатні працювати з нечіткою, неповною або частково достовірною інформацією. Суть нечіткої логіки полягає в апроксимації нелінійних залежностей за допомогою правил типу "ЯКЩО–ТО", які формуються на основі вхідних і вихідних даних. Для побудови таких правил із часових рядів необхідна навчальна вибірка, яка пов'язує m попередніх (вхідних) значень ряду з (m+1)-м (вихідним) значенням. Одним із найбільш популярних засобів реалізації цього підходу є ANFIS – адаптивна нейро-нечітка система виведення, що поєднує гнучкість нечітких правил із здатністю нейронних мереж до самонавчання. Розроблена дослідниками зі Стенфордського університету, система ANFIS набула широкого поширення завдяки своїй ефективності та доступності реалізації, зокрема у середовищі MATLAB [23]. Вона дозволяє автоматично формувати структуру правил, налаштовувати їх параметри та створювати прогностичні моделі, що здатні точно описувати складні нелінійні процеси. Таким чином, нейро-нечіткий підхід виявляється особливо цінним у випадках, коли впливові фактори важко формалізувати або описати традиційними математичними методами.

Упродовж останніх років зростає інтерес до глибоких нейронних мереж, таких як рекурентні нейронні мережі (RNN), їх модифікації на кшталт LSTM або GRU, які дозволяють ефективно враховувати послідовність спостережень та

довгострокові залежності в часових рядах. Завдяки здатності до навчання на великих масивах даних і гнучкості в адаптації до складних шаблонів, такі моделі дедалі частіше застосовуються в задачах прогнозування енергоспоживання, зокрема газу. Незважаючи на вищезазначене, протягом тривалого часу глибоке навчання поступалося класичним статистичним моделям у задачах прогнозування часових рядів. Це добре видно з результатів серії міжнародних змагань з прогнозування – так званих М-конкурсів Makridakis (M1, M2, M3, M4), які регулярно проводяться для оцінювання ефективності різних моделей на великій кількості реальних часових рядів [32]. До конкурсу M4 включно, найуспішніші моделі майже завжди базувалися на традиційних статистичних підходах або гібридних комбінаціях з елементами машинного навчання. У той же час методи, що повністю покладалися на глибоке навчання, демонстрували результати, що лише трохи перевищували базовий рівень конкурентів або й поступалися простішим статистичним алгоритмам. Проте ситуація почала змінюватися після публікації у 2020 році моделі N-BEATS – архітектури глибокого навчання, розробленої дослідниками компанії Element AI (Канада). Цей архітектурний підхід до глибокого навчання вперше продемонстрував перевагу над традиційними статистичними методами, випередивши переможця конкурсу M4. N-BEATS не потребує попередньої декомпозиції часового ряду, а замість цього автоматично навчається виявляти його внутрішню структуру, що значно спрощує процес побудови моделі та покращує її адаптивність до нових даних. Важливо зазначити, що архітектура N-BEATS була спеціально розроблена з урахуванням кількох ключових принципів: простоти, універсальності та здатності до інтерпретації. Модель не потребує використання спеціальних функцій або попереднього масштабування, що дозволяє зосередитися на потенціалі самої архітектури глибокого навчання без зовнішніх обмежень [31]. Крім того, модульна структура N-BEATS забезпечує можливість її подальшого розширення, що допомагає глибшому аналізу та інтерпретації отриманих результатів моделі та підвищує її практичну цінність у задачах прогнозування часових рядів.

Таким чином, вибір сучасного методу прогнозування обумовлюється низкою факторів, зокрема типом даних, ступенем їхньої складності, наявністю сезонних або трендових компонентів, а також вимогами до точності й інтерпретованості результатів. Успішність прогнозування значною мірою залежить від адекватності обраного методу до специфіки задачі та рівня підготовки вхідної інформації. Кожен із розглянутих підходів має свої переваги, що дозволяє підібрати інструмент, який найкраще відповідає особливостям конкретної задачі. Незважаючи на технологічне різноманіття, важливо не лише обрати відповідну модель, а й забезпечити коректну підготовку даних, перевірку її припущень та обґрунтовану оцінку результатів прогнозування.

Наступний розділ буде присвячено методичній частині дослідження – розгляду принципів реалізації обраних моделей, етапів їх побудови, вибору параметрів і підготовки даних для застосування в системі прогнозування.

1.3 Постановка задачі

В сучасних умовах зростаючої потреби в оптимізації використання енергоресурсів особливого значення набуває задача прогнозування споживання природного газу на рівні комунальних підприємств. Точне прогнозування дозволяє ефективніше планувати обсяги закупівель, оптимізувати фінансові витрати та підвищувати стабільність надання послуг споживачам.

В якості бази для побудови прогнозних моделей використовуються реальні дані про обсяги споживання газу комунальним підприємством за період з 2019 по 2025 рік. Дані отримані з відкритих джерел та представлені у табличному форматі із зазначенням обсягів споживання у помісячній роздільній здатності.

Перед початком моделювання здійснюється попередня обробка вхідних даних, яка включає:

- деталізацію обсягів споживання до подобового рівня для збільшення розмірності навчального набору та кращого виявлення сезонних закономірностей;

- обробку аномальних значень, що дозволяє зменшити вплив викидів і забезпечити більш стабільну якість прогнозування.

Для досягнення поставленої мети планується виконання таких основних етапів роботи:

- провести аналіз структури часового ряду: дослідити наявність трендів, сезонних компонент, перевірити стаціонарність ряду;
- сформулювати вимоги до моделей прогнозування з урахуванням виявлених властивостей даних;
- відібрати та побудувати кілька моделей прогнозування часових рядів, застосовуючи різні підходи (статистичні та інтелектуальні);
- виконати налаштування моделей та навчання на тренувальній вибірці;
- оцінити якість побудованих прогнозів за допомогою стандартних метрик прогнозування;
- побудувати графіки фактичних та прогнозованих значень для візуальної оцінки роботи моделей;
- провести порівняльний аналіз моделей та сформулювати рекомендації щодо вибору найбільш ефективного методу прогнозування;

Результатом роботи має стати:

- побудова прогнозних моделей споживання природного газу на основі реальних даних;
- порівняння ефективності різних методів прогнозування;
- формулювання рекомендацій щодо впровадження обраних моделей у практичну діяльність комунального підприємства для підвищення ефективності управління ресурсами.

РОЗДІЛ 2
МЕТОДИКА ПОБУДОВИ ТА ОЦІНКИ МОДЕЛЕЙ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

2.1 Архітектура та алгоритми прогнозних моделей

2.1.1 Моделі експоненційного згладжування

Методи експоненційного згладжування (exponential smoothing) належать до класичних статистичних підходів до прогнозування часових рядів. Ці методи здобули широке застосування завдяки своїй простоті, ефективності та здатності адаптуватися до змін у динаміці даних. Основна ідея полягає в тому, що нове значення прогнозу формується на основі попереднього прогнозу та поточного фактичного значення, при цьому більшій кількості останніх спостережень надається більша вага.

У загальному вигляді проста модель експоненційного згладжування (Simple Exponential Smoothing, SES) описується рівнянням:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \qquad (2.1)$$

де $\hat{y} + 1$ – прогнозоване значення на момент t+1;

y_t – фактичне значення на момент t;

$\hat{y}_t$ – прогнозоване значення на момент t;

$\alpha \in (0,1)$ – параметр згладжування, який визначає вплив поточного значення.

Ця модель добре працює для рядів без тренду та сезонності. Проте вона має обмеження у випадках, коли структура даних ускладнюється довготривалими змінами чи повторюваними сезонними коливаннями. Тому для більш складних структур були розроблені модифікації, які дозволяють враховувати трендову та сезонну компоненти.

Для рядів, що мають чітко виражену тенденцію (тренд), доцільно використовувати модель Хольта (Holt's Linear Trend Method), яка є розширенням SES. Вона вводить ще одну компоненту – трендову, яка оновлюється динамічно.

Рівняння моделі Хольта мають вигляд:

$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \quad (2.2)$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \quad (2.3)$$

$$\hat{y}_{t+h} = \ell_t + hb_t \quad (2.4)$$

де ℓ_t – рівень ряду на момент t;

b_t – тренд ряду на момент t;

α, β ∈ (0,1) – параметри згладжування для рівня та тренду відповідно;

$\hat{y}_{t+h}$– прогноз на h кроків уперед.

Для рядів, що демонструють як тренд, так і сезонні коливання, використовується модель Хольта–Вінтерса (Holt–Winters Method). Існує дві версії цієї моделі: адитивна та мультиплікативна. Адитивна версія підходить у разі стабільної амплітуди сезонних змін, а мультиплікативна – коли амплітуда сезонності зростає зі зростанням рівня ряду [15].

Адитивна модель Хольта–Вінтерса описується рівняннями:

$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \quad (2.5)$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \quad (2.6)$$

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \quad (2.7)$$

$$\hat{y}_{t+h} = \ell_t + hb_t + s_{t-m+h \bmod m} \quad (2.8)$$

де s_t – сезонна компонента;

m – довжина сезонного циклу (наприклад, 12 для місячних даних);

γ ∈ (0,1) – параметр згладжування сезонної компоненти;

решта позначень – як у моделі Хольта.

Ця модель дозволяє одночасно враховувати кілька важливих характеристик часового ряду й забезпечує досить точне прогнозування при регулярних сезонних коливаннях. Саме тому вона широко використовується в енергетичному секторі, включаючи задачі прогнозування споживання природного газу, де характерна яскраво виражена сезонність.

До недоліків моделей експоненційного згладжування можна віднести обмеженість у випадках складних нелінійних структур, високої нестабільності або відсутності чітко вираженого тренду. Також вони погано адаптуються до різких змін у структурі ряду, якщо ті не є регулярними.

Для побудови моделі експоненційного згладжування використовуватиметься бібліотека statsmodels у середовищі програмування Python.

2.1.2 Авторегресійні моделі

Авторегресійні моделі (AR) належать до класичних статистичних інструментів прогнозування часових рядів. Їх головна ідея полягає у припущенні, що поточне значення часового ряду можна описати як лінійну комбінацію попередніх значень ряду та випадкової похибки. Базова авторегресійна модель порядку p (позначається AR(p)) має вигляд:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \qquad (2.9)$$

де y_t – значення даних часового ряду в момент часу t;

c – константа (середній рівень);

φ_1,…, φ_p – параметри моделі (коефіцієнти авторегресії);

ε_t – білий шум (випадкова компонента з нульовим середнім та постійною дисперсією).

AR-моделі застосовуються до стаціонарних рядів, тобто таких, у яких середнє, дисперсія та автокореляції не змінюються з часом.

Модель ковзного середнього порядку q (MA(q)) описує поточне значення ряду як лінійну комбінацію випадкових похибок попередніх моментів часу:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (2.10)$$

де μ – середнє значення;

$\theta_1, \ldots, \theta_q$ – коефіцієнти ковзного середнього;

ε_t – білий шум.

MA-моделі також передбачають стаціонарність ряду.

Комбіновані моделі ARMA та ARIMA. Для більш повного опису властивостей стаціонарних рядів використовують модель ARMA(p, q), яка поєднує обидва підходи:

$$y_t = \mathrm{c} + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta \varepsilon_{t-j} \quad (2.11)$$

У випадку, якщо ряд не є стаціонарним, його необхідно диференціювати. Це призводить до моделі ARIMA (AutoRegressive Integrated Moving Average), яка має вигляд:

$$\Delta^d y_t = ARMA(p, q) \quad (2.12)$$

де $\Delta^d y_t = (1 - \mathrm{B})^d \mathrm{y}_t$ – диференційована версія ряду;

B – оператор зсуву: $By_t = y_{t-1}$;

d – порядок інтегрування (кількість диференціювань).

У випадках, коли ряд демонструє сезонні коливання, застосовують модель SARIMA (Seasonal ARIMA):

$$SARIMA(p, d, q)(P, D, Q)_s \quad (2.13)$$

де p, d, q – порядки авторегресії, диференціювання та ковзного середнього для несезонної частини;

P, D, Q – відповідні параметри для сезонної компоненти;

s – довжина сезонного циклу (наприклад, 12 для місячного ряду з річною сезонністю).

SARIMA дозволяє моделювати як короткострокові, так і довготривалі сезонні ефекти. Для оцінки параметрів зазвичай використовують максимізацію правдоподібності, а для вибору порядків – інформаційні критерії AIC або BIC.

Для попереднього визначення параметрів авторегресійних моделей, зокрема ARIMA та SARIMA, основними інструментами є графіки автокореляції (ACF) та часткової автокореляції (PACF). ACF дозволяє оцінити кількість лагів у компоненті ковзного середнього (параметр q), тоді як PACF допомагає визначити порядок авторегресії (p). Аналіз цих графіків дає змогу обґрунтовано підійти до вибору структури моделі ще до її навчання. Після вибору початкових значень параметрів моделі, її оптимізація здійснюється за критеріями AIC або BIC, що дозволяють оцінити баланс між точністю та складністю моделі.

Для реалізації авторегресійної моделі застосовуватиметься бібліотека statsmodels у Python.

Авторегресійні моделі та їх модифікації залишаються потужним інструментом для моделювання часових рядів. Основними перевагами авторегресійних моделей є чітка математична формалізація, наявність усталених процедур для підбору параметрів і здатність точно моделювати поведінку даних у разі стаціонарності. Проте вони мають і певні обмеження. Зокрема, ці моделі менш ефективні для даних зі складною нелінійною структурою або за наявності великої кількості зовнішніх факторів. Їх побудова потребує ретельної перевірки припущень щодо стаціонарності та нормального розподілу залишків.

2.1.3 Нечітко-хаотична модель

Нечітка логіка є потужним інструментом для моделювання нелінійних і складних систем, де традиційні методи моделювання можуть бути недостатніми. Особливо ефективним та поширеним прикладом її застосування є ANFIS (Adaptive Neuro-Fuzzy Inference System) – система виведення, що поєднує

нейронні мережі та нечітку логіку для адаптивного аналізу даних. Ця система базується на моделі нечіткого виведення типу Такагі–Сугено, а її архітектура дозволяє налаштовувати параметри моделі за допомогою методів навчання на основі даних.

Нечіткі апроксиматори, також відомі як нечіткі системи виведення, використовуються для моделювання залежностей «входи–вихід» на основі нечітких правил та функцій приналежності. Функції приналежності є ключовим елементом нечіткої логіки. Вони визначають ступінь приналежності певного значення до нечіткої множини. Існує кілька поширених типів функцій приналежності, кожен з яких має свої особливості та переваги:

Трикутні функції приналежності: Це одні з найпростіших функцій приналежності, які задаються трьома параметрами: a, b і c. Значення функції збільшується лінійно від 0 до 1 на інтервалі [a, b], потім лінійно зменшується від 1 до 0 на інтервалі [b, c]. Трикутні функції широко використовуються завдяки своїй простоті та ефективності.

Трапецієподібні функції приналежності: Вони задаються чотирма параметрами: a, b, c і d. Значення функції дорівнює 0 до a, лінійно збільшується від 0 до 1 на інтервалі [a, b], дорівнює 1 на інтервалі [b, c], а потім лінійно зменшується від 1 до 0 на інтервалі [c, d]. Трапецієподібні функції забезпечують більшу гнучкість порівняно з трикутними функціями.

Гауссівські функції приналежності: Вони мають форму дзвону і задаються двома параметрами: середнім значенням (c) і стандартним відхиленням (σ). Гауссівські функції часто використовуються для моделювання нечітких понять, що мають плавний перехід між належністю та неналежністю.

Узагальнені дзвоноподібні функції приналежності: Це узагальнена форма гауссівських функцій, яка дозволяє регулювати ступінь згладжування функції та її відхилення від симетричної форми. Вони задаються трьома параметрами: c (середнє значення), a (ширина) і b (нахил).

Сигмоїдальні функції приналежності: Вони мають S-подібну форму і часто використовуються для моделювання процесів, що мають поступовий перехід між

станами. Сигмоїдальні функції задаються двома параметрами, що визначають їх розташування та нахил.

Вибір типу функції приналежності залежить від конкретної задачі та характеристик даних, а також від способу представлення термів. Терми – це лінгвістичні змінні, що використовуються для опису нечітких множин. Наприклад, для опису температури можна використовувати терми "холодна", "прохолодна", "помірна", "тепла" та "гаряча". Кожен терм представлений певною функцією приналежності. Нечіткі правила є основою нечітких апроксиматорів. Вони складаються з передумови (нечіткі вхідні змінні) та висновку (нечітка вихідна змінна). Приклад нечіткого правила: "Якщо температура є холодною, а вологість є низькою, то опалення повинно бути високим".

Існує дві основні моделі нечітких апроксиматорів: Мамдані та Такагі-Сугено [29]. Модель Такагі-Сугено, що є основою ANFIS, використовує нечіткі множини як передумови правил, а висновки представлені лінійними або нелінійними функціями вхідних змінних. Ця модель є більш точною для апроксимації нелінійних залежностей і часто використовується в задачах регресії та прогнозування - саме тому на цій моделі буде зосереджено більше уваги, так як робота орієнтована саме на задачу прогнозування.

Принцип роботи нечіткого апроксиматора розглянуто на рис. 2.1 [9]:

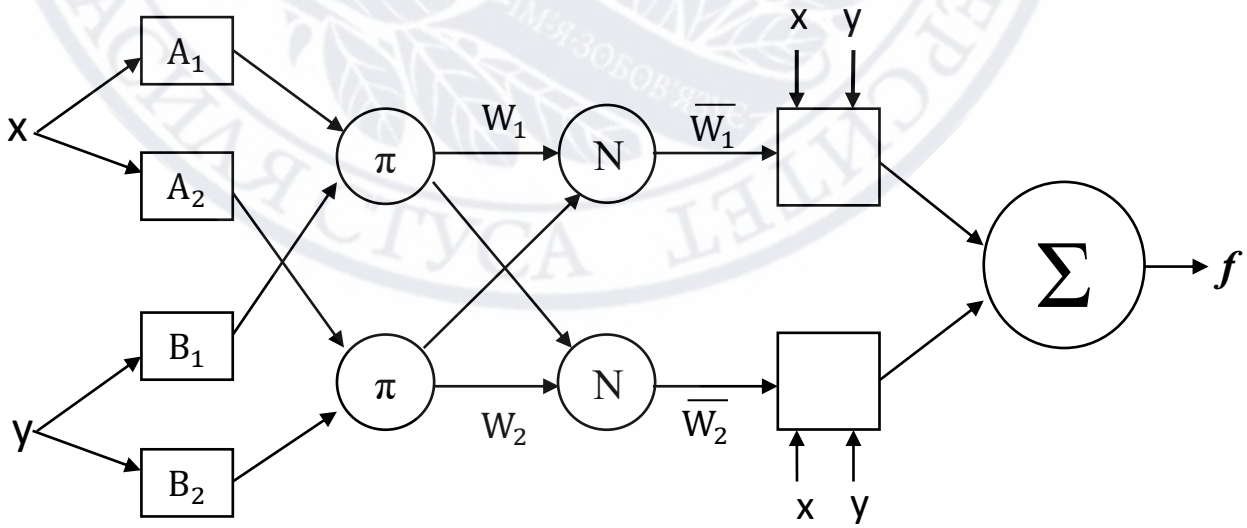

Рисунок 2.1 – Загальна структура моделі Такагі-Сугено ANFIS

На цьому рисунку зображена архітектура нечіткої системи виведення типу Такагі-Сугено. Вона складається з п'яти шарів:

Шар 1 (Layer 1): Це шар фазифікації, де вхідні змінні x та y перетворюються на ступені приналежності до відповідних нечітких множин A_i та B_i за допомогою функцій приналежності. Формула для визначення ступені приналежності наступна:

$$\mu_{ij}(x_i) = \frac{1}{1 - \left|\frac{x_i - c_{ij}}{a_{ij}}\right|^{2b_{ij}}} \tag{2.14}$$

де $\mu_{ij}(x_i)$ ступінь приналежності змінної x_i до терму A_{ij}, аналогічно змінної y_i до терму B_{ij};

a_{ij}, b_{ij}, c_{ij} – невідомі параметри.

Шар 2 (Layer 2): Шар правил, де обчислюються ступені активації кожного нечіткого правила шляхом множення ступенів приналежності передумов правила:

$$w_j = \prod_{i=1}^{n} \mu_{ij}(x_i) \tag{2.15}$$

де π - символ множення.

Шар 3 (Layer 3): Шар нормалізації, де ваги правил (ступені активації) нормалізуються діленням на суму ваг усіх правил (N - вузол нормалізації).

Шар 4 (Layer 4): Шар висновків, де обчислюються значення лінійних функцій f1, f2 тощо для кожного правила з нормалізованими вагами з попереднього шару.

Шар 5 (Layer 5): Шар композиції, де обчислюється зважена сума висновків усіх правил (∑ - вузол підсумовування), що є остаточним вихідним значенням системи:

$$f = \frac{\sum_{j=1}^{r} W_j f_i}{\sum_{j=1}^{r} w_J} \tag{2.16}$$

Така архітектура дозволяє моделі Такагі-Сугено ефективно апроксимувати нелінійні залежності між вхідними та вихідними змінними, використовуючи нечіткі правила з лінійними функціями висновків.

Модель ANFIS буде реалізована у середовищі MATLAB із використанням вбудованих інструментів для нейро-нечітких систем.

У роботі для налаштування нейро-нечіткої моделі ANFIS буде використано методи теорії хаосу, які дозволяють визначити оптимальні параметри часової затримки (τ) та розмірності простору (m). Вперше значимість інтеграції цих методів було описано і практично доведено в статті «Нечітко – хаотичні прогнозування часових рядів» у 2014 році, одним із її авторів є український вчений Ротштейн О. П. [38].

Параметри τ та m відіграють ключову роль у формуванні вхідного простору моделі та забезпеченні її здатності відтворювати динаміку часових рядів. Теоретичне обґрунтування вибору τ базується на методі взаємної інформації, який визначає ступінь залежності між значеннями ряду у моменти часу t та t+τ. Оптимальне значення τ відповідає першому локальному мінімуму функції взаємної інформації I(τ), як показано в роботах Фрейзера і Свінні [19].

Розмірність простору m обирається за допомогою методу хибних найближчих сусідів, доцільність його застосування вперше доведено в роботі [30], що дозволяє уникнути спотворень структури атрактора при відображенні його у фазовий простір. За цим підходом, фазовий портрет часової динаміки (графічне уявлення поведінки динамічної системи у просторі станів) відтворюється з одного скалярного ряду, і для цього будується векторний простір вбудування з параметрами τ і m, де кожен вектор представляє стан системи у заданий момент.

Обчислення методів для підбору параметрів τ та m буде здійснюватися за допомогою методів взаємної інформації та хибних найближчих сусідів вбудованих в програмне забезпечення Matjaž Perc пакету Nonlinear time series analysis [31].

Для застосування методів дані потрібно попередньо підготувати, нормувавши часовий ряд в інтервалі [0, 1] за наступною формулою:

$$\tilde{x}_i = \frac{x_i - \underline{x}}{\bar{x} - \underline{x}}, \tag{2.17}$$

де $\tilde{x}_i$ – нормоване значення;

$\bar{x}(\underline{x})$ – максимальне (мінімальне) значення.

2.1.4 Глибока нейромережева модель

Модель N-BEATS є прикладом сучасного підходу до прогнозування часових рядів на основі архітектури глибокого навчання, що не вимагає попередньої декомпозиції чи аналізу статистичних властивостей ряду. Основною ідеєю моделі є побудова прогнозу шляхом послідовного додавання прогнозних блоків, які працюють у зворотному порядку (backcast) для реконструкції вхідного сигналу та у прямому напрямку (forecast) для формування прогнозу.

N-BEATS має стекову архітектуру, де кожен стек складається з кількох блоків, тоді як кожен блок самостійно генерує частковий прогноз, а також реконструює частину вхідного сигналу, що дозволяє поступово уточнювати прогноз на наступних кроках. В основі кожного блоку лежить багаторівнева повнозв'язна нейронна мережа з функцією активації ReLU [31]. Вхідні дані проходять через кілька прихованих шарів, після чого розділяються на два виходи:

$$\hat{x}_\ell = \sum_{i=1}^{\dim(\theta^b)} \theta^b_{\ell,i} v^b_i, \quad \hat{y}_\ell = \sum_{i=1}^{\dim(\theta^f)} \theta^f_{\ell,i} v^f_i \tag{2.18}$$

де θ^b, θ^b коефіцієнти проєкції у зворотній та прямій гілках відповідно;

v^b_i, v^f_i базисні вектори.

Базисні вектори можуть бути різних типів, зокрема:

Поліноміальні функції – використовуються для моделювання тренду, відображають зростаючу або спадну динаміку.

Гармонійні функції – синусоїдальні та косинусоїдальні складові, ефективні для сезонних і циклічних патернів.

Параметризовані функції – виводяться автоматично під час тренування, не мають заданої форми і забезпечують додаткову гнучкість, особливо в generic-архітектурі, яку буде розглянуто далі.

Кожен блок приймає на вхід залишковий сигнал , обчислює його backcast, після чого сигнал передається далі у вигляді різниці:

$$x_{\ell+1} = x_\ell - \hat{x}_\ell \tag{2.19}$$

Паралельно блокується частковий прогноз, який потім додається до сумарного прогнозу:

$$\hat{y} = \sum_{\ell} \hat{y}\ell \tag{2.20}$$

Це реалізує принцип подвійного залишкового складання (doubly residual stacking), що дозволяє не лише послідовно зменшувати помилку реконструкції, а й робити остаточний прогноз як суму прогнозів окремих блоків, що підвищує стійкість та точність моделі.

Архітектурно модель складається з кількох стеків, кожен з яких містить набір послідовно з’єднаних блоків. Після обробки даних одним стеком, залишок backcast передається у наступний стек. Кожен стек формує частковий прогноз, який агрегується на фінальному етапі. Така ієрархічна структура дозволяє створити дуже глибоку архітектуру без втрати стабільності при навчанні. Візуалізація архітектури моделі наведена на рис. 2.2.

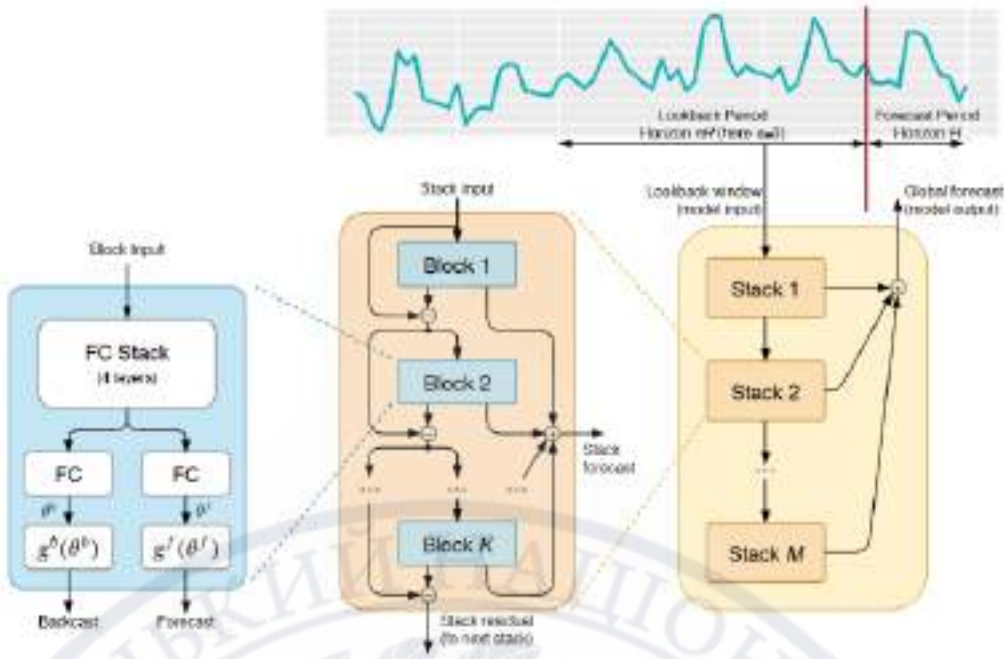

Рисунок 2.2 – Архітектура моделі N-BEATS в межах одного стеку

Тут зображено послідовну побудову прогнозу у межах одного стеку: кожен блок виконує одночасно два завдання – прогнозує майбутні значення (forecast) і відновлює частину історії (backcast). Backcast-вихід використовується для обчислення залишку, який подається на вхід наступного блоку, а часткові прогнози підсумовуються для отримання остаточного результату. Такий механізм дозволяє моделі поетапно уточнювати прогноз, зберігаючи можливість аналізу внеску окремих блоків у фінальний результат. У моделі також передбачено два підходи до побудови архітектури стеків – загальний (generic) і інтерпретований (interpretable) [31]. На рис. 2.3 представлено графічне порівняння цих двох підходів.

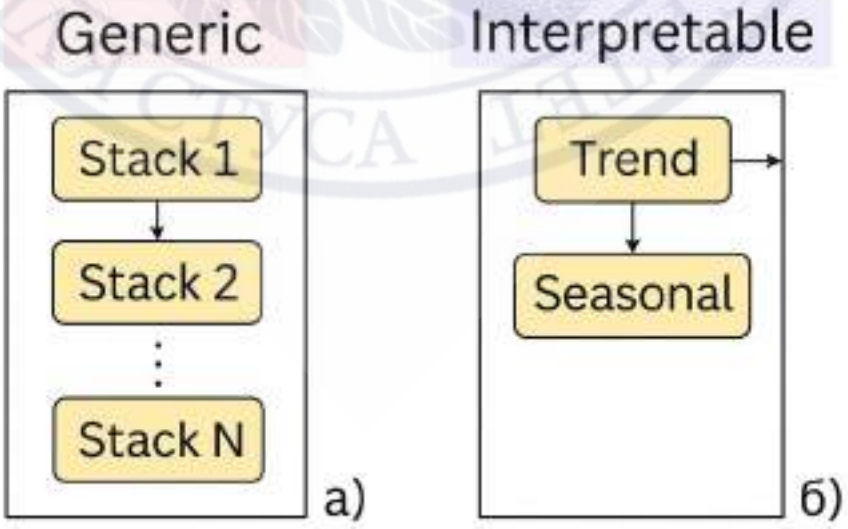

Рисунок 2.3 Загальний (а) та інтерпретований (б) підходи до побудови архітектури стеків

У загальному варіанті стеків (рис. 2.3 (а)) модель складається з набору стеків, що по черзі обробляють залишковий сигнал, при цьому форма базисних функцій не обмежується. У випадку інтерпретованої архітектури (рис.2.3 (б)) кожен стек відповідає за окремий компонент часового ряду – тренд або сезонність – що дозволяє зробити вихід моделі більш прозорим для користувача.

У випадку інтерпретованої архітектури для трендової складової використовуються поліноміальні функції:

$$\hat{y}_{s,l}^{tr} = \sum_{i=0}^{p} \theta_{s,\ell,i}^{f} t^{i} \tag{2.21}$$

де $\theta_{s,\ell,i}^{f}$, – коефіцієнти полінома, що навчаються, а t – часова координата у вікні прогнозу.

Для моделювання сезонності застосовується розклад у ряди Фур'є

$$\hat{y}_{s,l}^{seas} = \sum_{i=0}^{[H/2]-1} \theta_{s,\ell,i}^{f} \cos(2\pi t) + \theta_{s,\ell,i+[H/2]}^{f} \sin(2\pi t) \tag{2.22}$$

де H – горизонт прогнозу. Це дозволяє моделі точно відтворювати циклічні коливання, характерні для сезонних патернів у часових рядах.

У моделі також передбачено використання спільних базисних векторів у межах одного стеку, що покращує інтерпретованість та стабільність. Для кожного forecast і backcast виходу передбачено власний набір параметрів, що дозволяє моделі зберігати гнучкість у представленні залежностей.

Завдяки описаній архітектурі модель N-BEATS здатна:

- працювати без попередньої декомпозиції часового ряду;
- автоматично виявляти складні тренди та сезонні патерни;
- зберігати прозорість та інтерпретованість за рахунок обмежених базисів (у випадку інтерпретованих блоків);
- масштабуватись до великої кількості часових рядів або довгих періодів прогнозу.

Для побудови моделі N-BEATS використовуватиметься бібліотека NeuralForecast у середовищі Python, яка забезпечує гнучке налаштування архітектури глибинних нейронних мереж для прогнозування часових рядів.

2.1.5 Базова модель

Крім складних моделей прогнозування, у роботі також використовуватиметься базова модель (baseline), яка слугує точкою порівняння для оцінки точності. Як базову обрано модель «наївного прогнозу», що передбачає повторення останнього наявного значення ряду як прогнозного. Незважаючи на простоту, така модель дозволяє виявити, наскільки складніші алгоритми забезпечують покращення результатів і чи є доцільним їх використання з урахуванням витрат на реалізацію.

У наступному підпункті буде розглянуто підхід до підготовки даних для тренування моделей прогнозування, зокрема побудову вибірки, формування вхідних векторів, нормалізацію та валідацію.

2.2 Підготовка даних для прогнозування часових рядів

2.2.1 Попередній аналіз і обробка даних

Першим етапом підготовки даних є їх попередній аналіз та обробка, що має на меті перевірку якості вхідного часового ряду, виявлення проблем та приведення даних до стану, придатного для моделювання. Зокрема, необхідно виконати:

- перевірку на пропущені значення. Якщо у часовому ряду наявні пропуски, їх потрібно або заповнити відповідним методом (лінійна інтерполяція, forward/backward fill, середнє значення) або вилучити ці спостереження, якщо вони поодинокі та не впливають на загальну динаміку. У цій роботі використовуються дані без пропусків, тому потреба в цьому кроці відсутня.
- виявлення викидів. Аномальні значення можуть бути результатом технічних помилок або нетипових подій. Вони виявляються за допомогою статистичних критеріїв, таких як міжквартильний розмах (IQR), Z-оцінка

або візуальний аналіз. Зокрема, метод Z-оцінки є ефективним у випадках великих вибірок (понад тисячу спостережень), оскільки забезпечує достовірне оцінювання середнього значення та стандартного відхилення. Якщо розподіл значень приблизно симетричний, то Z-метод дозволяє надійно виявити екстремальні відхилення, які можуть спотворювати результати моделювання;

- перевірку стаціонарності. Оскільки частина моделей вимагає стаціонарного ряду, необхідно виявити наявність тренду або сезонних коливань. Для цього використовуються графічні методи – зокрема візуалізація самого часового ряду та аналіз автокореляційних графіків (ACF та PACF). Такий підхід дозволяє оцінити характер змінності даних, виявити тренди або повторювані сезонні структури, що свідчать про нестаціонарність. У разі її виявлення ряд трансформується шляхом диференціювання.

Ці дії формують основу, що забезпечує якісну та послідовну роботу всіх етапів моделювання.

2.2.2 Агрегація і розгортання даних

Часові ряди можуть надаватися з різною частотою – від щохвилинної до щомісячної. У реальних умовах часто доводиться адаптувати такі дані під потреби моделі:

Агрегація даних передбачає об'єднання показників за більший інтервал часу. Наприклад, з поденних даних можна формувати тижневі або місячні суми/середні. Це доцільно, коли прогноз будується на тривалий період, або коли потрібно зменшити шум у даних.

Розгортання даних виконується у протилежному випадку – коли є потреба перетворити дані нижчої частоти (наприклад, помісячні) у більш деталізовані (поденні), щоб забезпечити більшу кількість точок навчальної вибірки. У даній роботі, з метою збільшення розмірності навчального набору, було виконано умовне розгортання помісячних даних у поденні шляхом розподілу обсягу

споживання в межах кожного місяця. Це дозволяє моделі краще захопити тренди та сезонність, підвищити узагальнюючу здатність, скоротити ризик перенавчання – ситуації, коли модель занадто точно відображає навчальні дані, але втрачає здатність узагальнювати на нових, – а також провести валідацію - перевірку моделі на відкладеній частині даних з навчального набору, та тестування - оцінювання точності на зовсім нових даних.

2.2.3 Формування вибірки для навчання та тестування

У цьому підпункті розглядається формування вибірки для різних типів моделей, які використовуються в роботі.

Для моделей ARIMA та експоненціального згладжування навчання здійснюється на повному часовому ряду без формування окремих навчальних прикладів у форматі «вхід–вихід». У таких моделях прогноз формується на основі попередніх значень ряду.

Поділ на навчальну та тестову вибірки виконується шляхом відокремлення останньої частини часового ряду (10–20%) для тестування, тоді як решта даних використовується для навчання моделі. Такий підхід дозволяє імітувати реальні умови прогнозування, коли майбутні значення ще невідомі, а моделі навчаються на історичних даних.

Для моделей ANFIS та N-BEATS формування навчальної вибірки потребує створення прикладів у форматі «вхід → вихід». Наприклад, для моделі N-BEATS:

- вхід (X) – послідовність довжини input_size (кількість попередніх точок часового ряду, які модель використовує для формування прогнозу);
- вихід (y) – прогнозовані значення на обраний інтервал (параметр h (horizon)).

Аналогічно формується вибірка для ANFIS, де структура вхідних векторів базується на обраних параметрах часової затримки (τ) та розмірності простору (m).

У підсумку з усього часового ряду формується відповідна кількість навчальних прикладів. Надалі ці приклади поділяються на:

- навчальну вибірку – ~80% прикладів для тренування моделі;
- валідаційну вибірку – ~10% для контролю навчання, в разі необхідності;
- тестову вибірку – ~10% для фінального оцінювання якості моделі.

Подібний поділ забезпечує контроль якості навчання, запобігає перенавчанню та дозволяє перевірити здатність моделі узагальнювати нову інформацію.

2.3 Оцінка точності та адекватності моделей прогнозування

2.3.1 Метрики точності прогнозу

Для оцінки точності моделей у даному дослідженні використовуються три ключові метрики: MAE – середня абсолютна помилка, RMSE – середньоквадратична помилка, sMAPE – симетризована середня абсолютна відносна помилка та MASE – середня абсолютна шкалована помилка. Кожна з них дає змогу з різних боків оцінити відповідність моделі до реальних даних.

Середня абсолютна помилка:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{2.23}$$

де $\hat{y}_i$ – прогнозоване значення;

y_i – фактичне значення;

n – кількість точок у вибірці.

Середньоквадратична помилка:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{2.24}$$

де $\hat{y}_i$ – прогнозоване значення;

y_i – фактичне значення;

n – кількість точок у вибірці.

RMSE вимірює середнє квадратичне відхилення прогнозованих значень від фактичних. Вона є чутливою до великих помилок, тому особливо корисна, коли потрібно виявити суттєві відхилення прогнозу.

Симетризована середня абсолютна відносна помилка:

$$sMAPE = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{(|y_i| + |\hat{y}_i|) \div 2} \tag{2.25}$$

де $\hat{y}_i$ – прогнозоване значення;

y_i – фактичне значення;

n – кількість точок у вибірці.

sMAPE усуває асиметрію MAPE та гарантує значення у межах [0%; 200%], що робить її придатнішою для порівняння моделей у задачах з різними масштабами даних.

Середня абсолютна шкалована помилка:

$$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_i - i|} \tag{2.26}$$

де $\hat{y}_i$ – прогнозоване значення;

y_i – фактичне значення;

n – кількість точок у вибірці;

знаменник – середня абсолютна помилка наївної моделі з лагом 1.

MASE є безрозмірною метрикою, що масштабує абсолютну помилку на основі середньої помилки наївної моделі (з лагом 1). Значення MASE < 1 вказує, що модель є точнішою за наївний прогноз, MASE > 1 – навпаки. Метрика MASE дозволяє безпосередньо порівнювати точність прогнозної моделі з базовою моделлю – у даному випадку з наївною, яка повторює останнє відоме значення. Завдяки цьому можна кількісно оцінити, наскільки складніші методи перевершують простий підхід.

Усі ці метрики будуть використані для комплексного оцінювання якості прогнозів моделей, що дозволяє врахувати як абсолютну, так і відносну точність, а також порівняти різні підходи до прогнозування.

2.3.2 Метрики якості побудованої моделі

Для оцінки того, наскільки модель відповідає даним, особливо в класичних підходах на зразок ARIMA або моделей експоненційного згладжування, використовуються специфічні статистичні метрики:

AIC (Akaike Information Criterion): показник, що враховує якість апроксимації моделі до даних із урахуванням кількості параметрів.

$$AIC = 2k - 2ln(L) \tag{2.27}$$

де k – кількість параметрів моделі;

L – максимальне значення функції правдоподібності.

Чим нижче значення AIC, тим кращою вважається модель.

BIC (Bayesian Information Criterion): подібний до AIC, але вводить жорсткіше штрафування за складність моделі.

$$BIC = kln(n) - 2ln(L) \tag{2.28}$$

де n – кількість спостережень.

Використовується для відбору більш компактних моделей за умови великої кількості даних.

Тест Льюнга–Бокса (Ljung–Box test): статистичний тест для перевірки автокореляції залишків моделі. Нульова гіпотеза тесту: залишки є незалежними (відсутність автокореляції). Статистика тесту:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{p}_k^2}{n-k} \tag{2.29}$$

де n – кількість спостережень;

h – кількість лагів;

$\hat{p}_k$ – автокореляція залишків на лагу k.

2.3.3 Візуальні засоби оцінки якості прогнозу

Важливо не лише аналізувати числові метрики, але й перевіряти візуальні узгодження між прогнозом і реальними даними. Серед основних підходів, які будуть використані:

Графік реального ряду та прогнозу – дозволяє побачити наскільки добре модель передбачає тренди, сезонність, зміни рівня.

ACF/PACF залишків – автокореляційна та часткова автокореляційна функції залишків, які мають бути близькі до випадкових, якщо модель адекватна.

Графік щільності розподілу залишків – дозволяє оцінити, чи залишки приблизно нормально розподілені, що часто є припущенням класичних моделей.

Графік залишків (residuals) – аналіз поведінки залишків (відхилень між прогнозом і фактичним значенням), що дозволяє виявити систематичні помилки моделі.

РОЗДІЛ 3
ПРАКТИЧНА РЕАЛІЗАЦІЯ

У процесі практичної реалізації використовувалися два середовища програмування. Основна частина моделювання часових рядів виконувалась у середовищі Google Colab із використанням мови програмування Python, що забезпечувало гнучкість у реалізації алгоритмів та ефективну обробку великих обсягів даних. Для розробки та навчання моделі ANFIS було застосовано середовище MATLAB, яке має вбудовані засоби для створення нейро-нечітких систем та полегшує роботу з нечіткою логікою та регресійним аналізом.

3.1 Аналіз та обробка вхідних даних

Для побудови прогнозних моделей було використано реальні дані споживання комунальних ресурсів КП "Центральний міський стадіон" м. Вінниці. Первинні дані були отримані з відкритого доступу на порталі Дія [2] і охоплювали період з 2019 по березень 2025 року із місячною частотою спостережень. Початково дані були сформовані окремими файлами за кожен календарний рік, а структура кожного файлу включала такі стовпці: id, date, organizationName, organizationID, providerName, providerID, meterNumber, type, quantity, unitName.

Оскільки вихідні дані містили інформацію про споживання різних ресурсів, для подальшого аналізу було відібрано лише записи, що стосуються споживання природного газу. Було залишено тільки необхідні стовпці: дату спостереження (date) та кількість спожитого газу (quantity), виражену в кубічних метрах.

Для збільшення розмірності навчального набору було виконано розгортання помісячних даних у подобові шляхом розподілу обсягу споживання всередині кожного місяця. На рис. 3.1 представлено часовий ряд споживання газу, підготовлений для подальшого аналізу та моделювання.

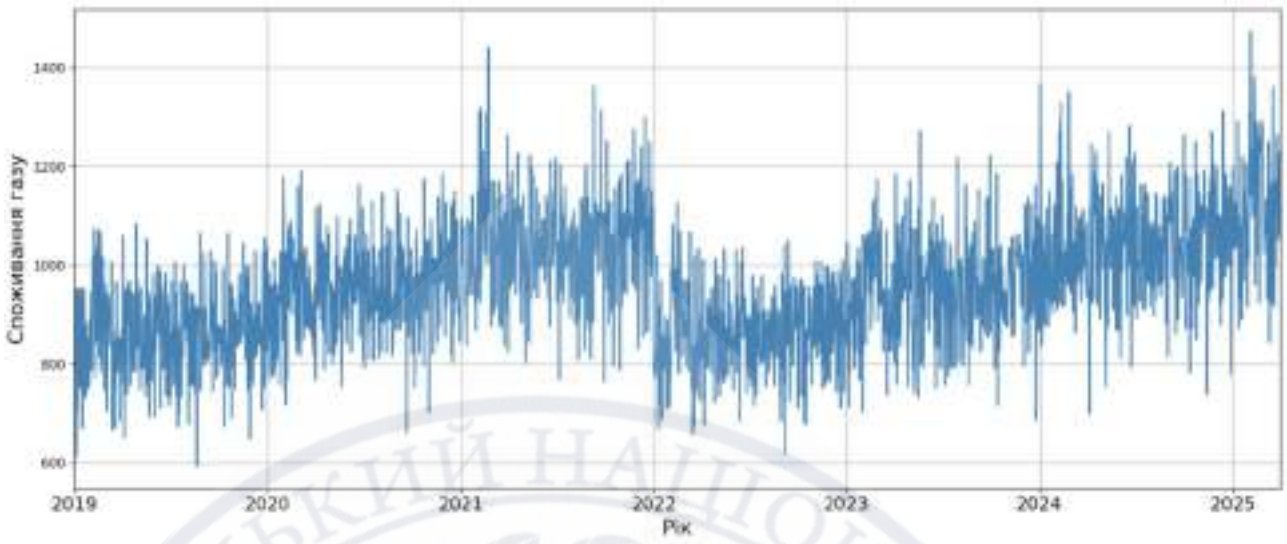

Рисунок 3.1 – Подобовий ряд споживання газу КП "Центральний міський стадіон"

На рис. 3.2 наведено розподіл обсягів споживання газу за окремими роками. З графіка видно загальну тенденцію до коливань обсягів у різні роки, із помітним зниженням споживання у 2022 році. Такий спад пов'язаний із зовнішніми чинниками, зокрема впливом повномасштабної війни в Україні, що призвело до скорочення обсягів споживання у багатьох секторах економіки.

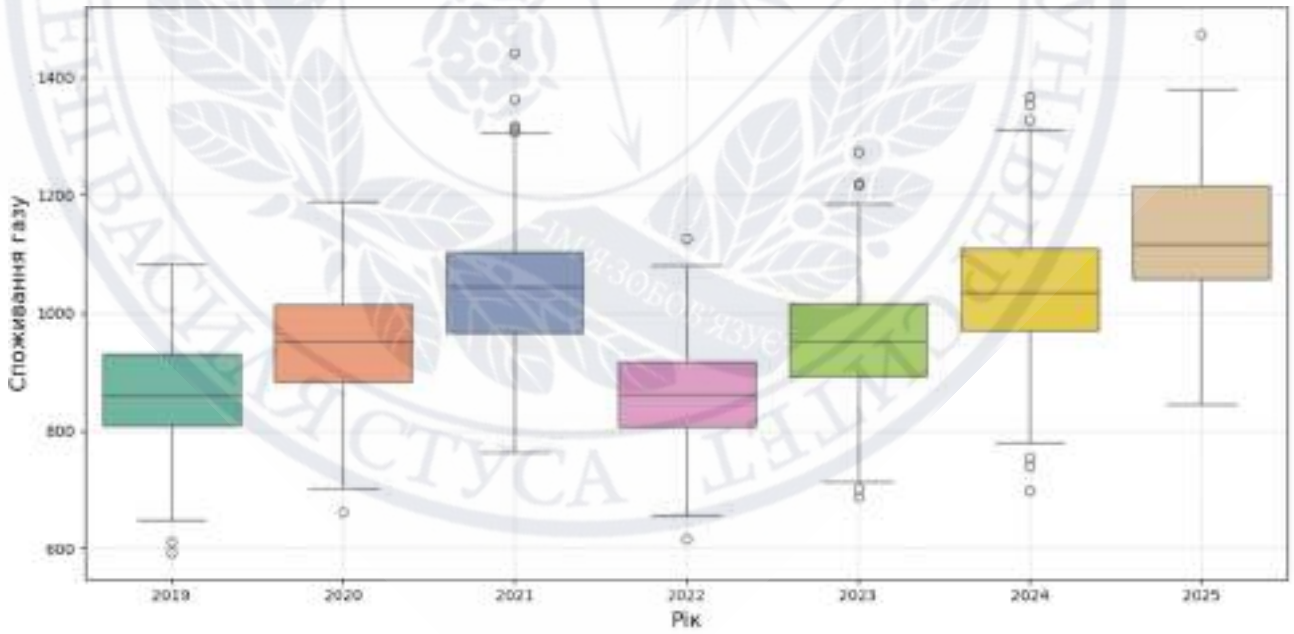

Рисунок 3.2 – Розподіл подобового споживання газу за роками

3.1.1 Виявлення та обробка аномальних значень

Для попереднього очищення даних було виконано виявлення аномальних значень за допомогою методу Z-оцінки (рис. 3.3).

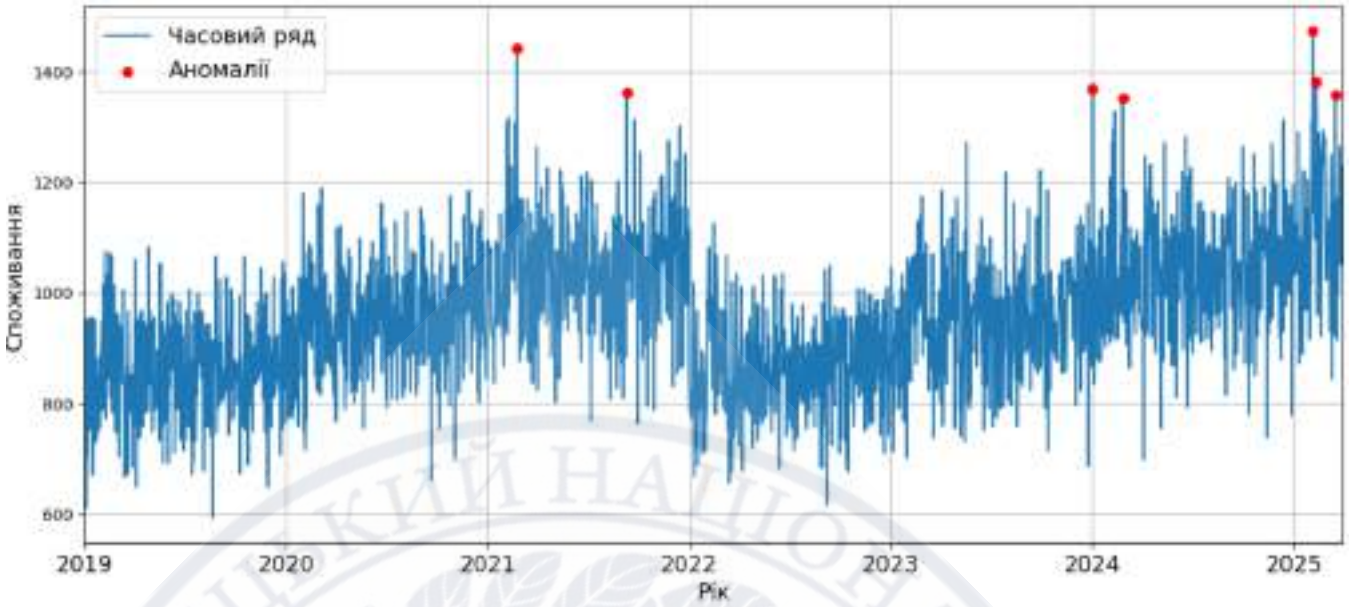

Рисунок 3.3 – Аномальні значення даних виявлені за допомогою Z-оцінки

Усі точки, що перевищували три стандартні відхилення від середнього, були позначені як потенційні викиди та замінені шляхом лінійної інтерполяції (рис. 3.4).

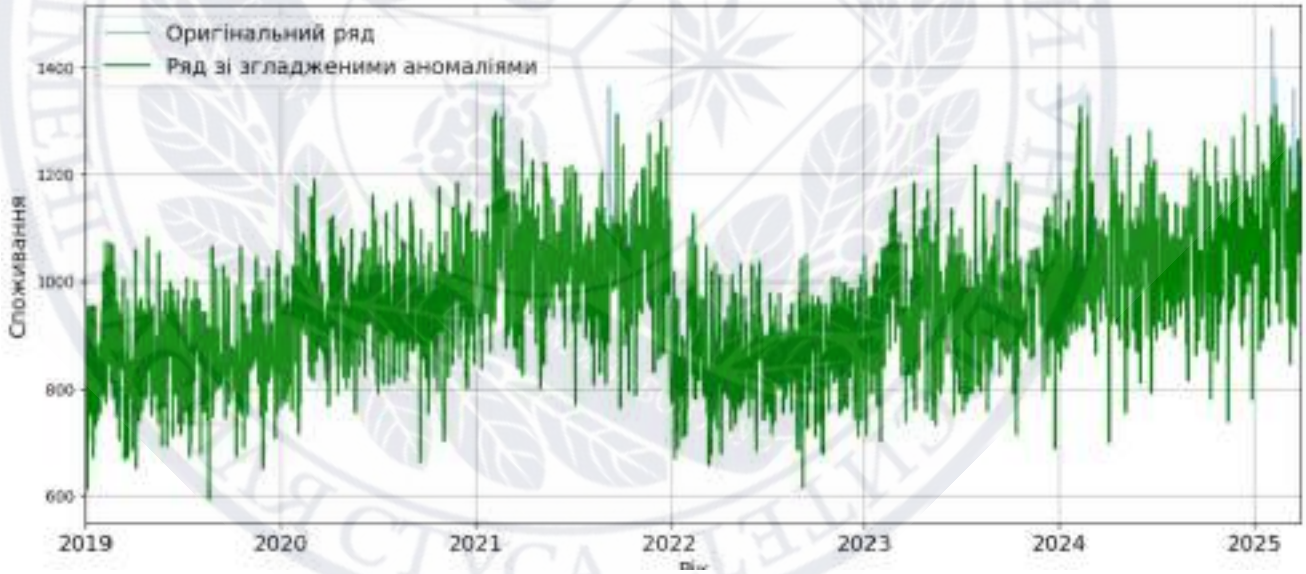

Рисунок 3.4 – Вигляд часового ряду після згладжування аномальних значень

Такий підхід дозволив зберегти загальну структуру ряду, усунувши вплив нетипових сплесків перед побудовою моделей прогнозування.

3.1.2 Декомпозиція часового ряду

Для кращого розуміння структури часового ряду споживання газу було виконано його адитивну декомпозицію. Оскільки амплітуда сезонних коливань залишалася сталою впродовж усього періоду спостереження це вказує на доцільність використання саме адитивної моделі. На рис. 3.5 представлено результати декомпозиції.

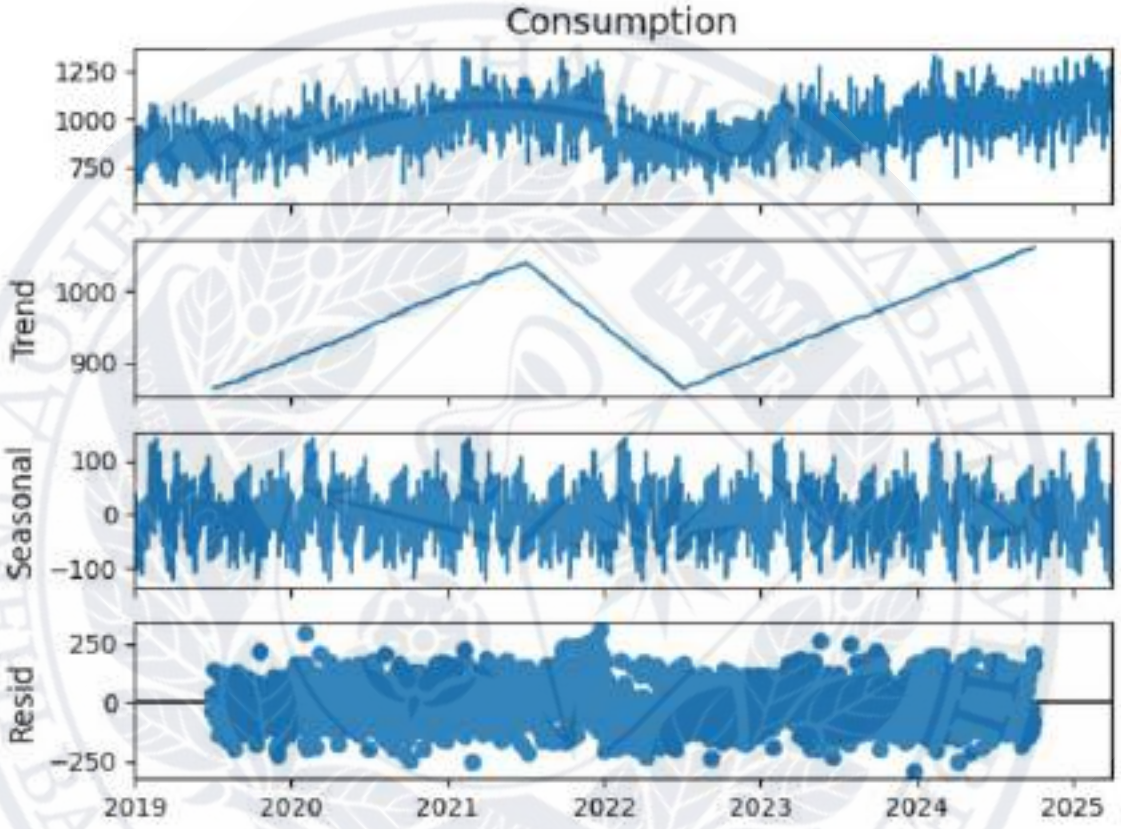

Рисунок 3.5 – Адитивна декомпозиція часового ряду споживання газу

Аналіз декомпозиції показує наступне:

- трендова складова має виражену зміну: спочатку зростання, потім спад, після чого знову зростання, що може свідчити про вплив зовнішніх факторів, наприклад сезонів чи змін у споживанні підприємством;
- сезонна складова демонструє регулярні пікові значення, характерні для зимового періоду, ймовірно, через опалення;

- залишкова складова є досить варіативною, без яскраво вираженої автокореляції, що вказує на адекватність виокремлення трендової та сезонної компонент.

Результати декомпозиції часового ряду демонструють його складну структуру з вираженими трендовими змінами та сезонними коливаннями, що свідчить про його нестаціонарність. Це означає, що дані потребують додаткової підготовки перед побудовою прогнозної моделі.

3.2 Налаштування та реалізація моделей

3.2.1 Побудова моделі експотенційного зглажування

Для моделювання та прогнозування споживання газу було обрано модель експоненційного згладжування Хольта–Вінтерса, яка враховує трендову та сезонну компоненти часового ряду. Вибір обумовлений результатами попереднього аналізу: декомпозиція виявила змінний тренд та стабільну сезонність адитивної природи (рисунок 3.X).

У дослідженні було побудовано моделі з такими параметрами:

- trend='add' – адитивний тренд через сталу швидкість зміни рівня ряду;
- seasonal='add' – адитивна сезонність через стабільну амплітуду коливань;
- seasonal_periods=30 або 365 – для перевірки впливу різних періодичностей.

Хоча основний аналіз зосереджувався на адитивних моделях, для формального порівняння також протестовано мультиплікативні варіанти. Загалом було реалізовано чотири моделі. Для уникнення надмірного зростання або зниження тренду в довгостроковій перспективі застосовано загасаючий тренд (damped_trend=True), що зробило прогноз більш стабільним та реалістичним для задач споживання ресурсів.

Порівняння моделей показало, що найкращі значення AIC та BIC має модель з адитивними компонентами та сезонністю 30 днів (AIC = 17960.14, BIC = 18155.23). Це вказує на оптимальний баланс між точністю апроксимації та складністю.

На рис. 3.6 наведено результати навчання обраної моделі експоненційного згладжування Хольта–Вінтерса. Оптимізовані значення коефіцієнтів згладжування рівня (α = 0.0866), тренду (β = 0.0001) та сезонності (γ = 0.0671) свідчать про помірну адаптивність моделі до змін у часовому ряді, що відповідає природі досліджуваних даних.

Окрім того, у таблиці наведено початкові оцінки рівня, тренду та сезонних складових, що були автоматично визначені під час навчання.

ExponentialSmoothing Model Results

Dep. Variable:	Consum	**No. Observations:**	1947
Model:	ExponentialSmoothing	**SSE**	19050645.060
Optimized:	True	**AIC**	17960.139
Trend:	Additive	**BIC**	18155.231
Seasonal:	Additive	**AICC**	17961.612
Seasonal Periods:	30	**Date:**	Wed, 07 May 2025
Box-Cox:	False	**Time:**	21:40:22
Box-Cox Coeff.:	None		

	coeff	code	optimized
smoothing_level	0.0866470	alpha	True
smoothing_trend	0.0001316	beta	True
smoothing_seasonal	0.0671188	gamma	True
initial_level	805.13827	l.0	True
initial_trend	1.4077563	b.0	True
damping_trend	0.9877493	phi	True
initial_seasons.0	27.173177	s.0	True
initial_seasons.1	21.938526	s.1	True
initial_seasons.2	-18.898336	s.2	True
initial_seasons.3	-6.7943384	s.3	True

Рисунок 3.6 – Фрагмент результатів навчання моделі Holt–Winters

Для кількісної оцінки залишкової автокореляції у моделі Хольта–Вінтерса застосовано тест Льюнга–Бокса за допомогою функції acorr_ljungbox з бібліотеки statsmodels.stats.diagnostic. За результатами тесту при лагах 10, 20 та 30 (p-значення 0.99, 0.94 та 0.60 відповідно), автокореляція залишків не виявлена, оскільки всі p-значення перевищують критичний рівень 0.05. Це підтверджує статистичну адекватність моделі.

Для перевірки нормальності розподілу залишків побудовано гістограму з накладеною теоретичною кривою нормального розподілу. З рисунка 3.7 видно, що залишки мають симетричну форму з концентрацією навколо нуля, що

свідчить про наближення до нормального розподілу і дозволяє довіряти побудованим прогнозним інтервалам.

Рисунок 3.7 – Гістограма розподілу залишків моделі Хольта–Вінтерса з накладеною кривою нормального розподілу

На основі обраної моделі Хольта–Вінтерса було здійснено прогнозування споживання газу на річний тестовий період (з квітня 2024 по березень 2025 року). На рис. 3.8 представлено порівняння прогнозних значень із фактичними даними. Як видно з графіка, модель здебільшого адекватно відтворює загальний рівень та сезонну структуру споживання, хоча в окремі моменти спостерігаються відхилення, що можуть бути зумовлені випадковими коливаннями або обмеженнями самої моделі у відтворенні короткострокової динаміки.

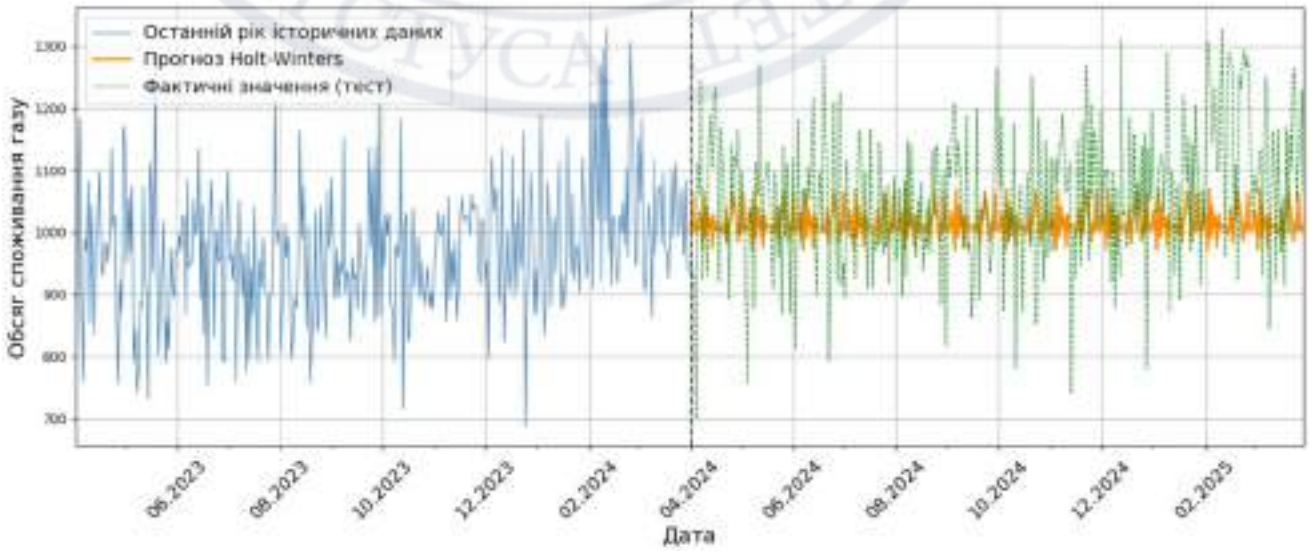

Рисунок 3.8 – Прогноз споживання газу на тестових даних моделлю Holt-Winters

Результати кількісного оцінювання прогнозної якості представлені в таблиці 3.1.

Таблиця 3.1 Результати оцінювання точності прогнозу моделі Holt-Winters

Метрика	Тренувальні дані	Тестові дані
MAE	78.79	97.69
RMSE	98.92	122.05
sMAPE (%)	8.42	9.30
MASE	0.76	0.90
AIC	17960.14	
BIC	18155.23	

Аналіз показників точності моделі Хольта–Вінтерса свідчить про її стабільну роботу як на тренувальних, так і на тестових даних. Значення MAE та RMSE залишаються на порівнянному рівні, без суттєвого зростання похибки на тестовому періоді, що свідчить про відсутність перенавчання. Значення sMAPE для обох вибірок майже однакове, що підтверджує узгодженість моделі у відносному вимірі. Показник MASE також залишається меншим за одиницю на обох етапах, що означає перевагу моделі над наївним прогнозом. Таким чином, модель демонструє прийнятну якість прогнозування для подальшого практичного використання.

3.2.2 Побудова авторегресійсної моделі

Після проведеної декомпозиції часового ряду було виявлено наявність вираженого тренду та сезонної складової, що свідчить про складну структуру даних. Оскільки класичні авторегресійні моделі, не враховують сезонності без додаткових перетворень, було прийнято рішення використати розширену модель SARIMA, яка дозволяє враховувати як сезонні ефекти, так і інші структурні компоненти ряду.

Для налаштування параметрів цієї моделі необхідно проаналізувати автокореляцію часового ряду. Для цього було побудовано графіки ACF - автокореляційна функція та PACF - часткова автокореляційна функція (рис. 3.9). Перед побудовою ACF та PACF функцій часовий ряд було приведено до стаціонарного вигляду шляхом першого диференціювання, оскільки попередній аналіз (рис. 3.9) виявив наявність чітко вираженого тренду.

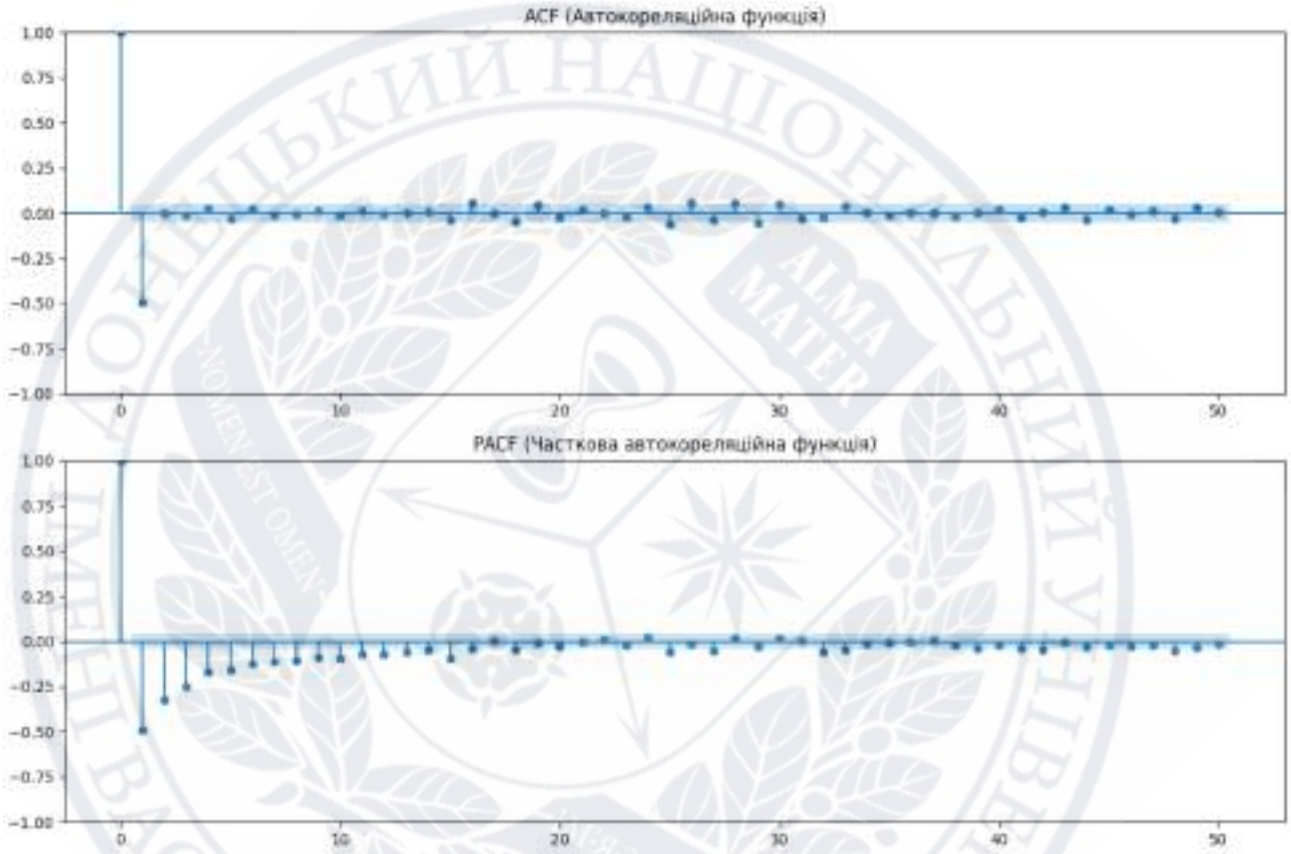

Рисунок 3.9 – Графіки ACF та PACF функцій після диференціювання ряду

Аналіз графіків автокореляційної та часткової автокореляційної функцій після першого диференціювання показав, що часовий ряд набув стаціонарного вигляду: ACF не демонструє повільного спадання, а PACF обмежується кількома лагами. Це свідчить про відсутність довгострокових залежностей та обґрунтовує вибір параметрів d=1, D=1, що відповідає наявності тренду та умовної сезонності з періодом близько 30 днів. Повільне згасання ACF вказує на необхідність включення компоненти ковзного середнього (q=1), тоді як обмеження PACF після першого лага підтверджує доцільність авторегресійної компоненти (p=1).

Незважаючи на слабше виражені сезонні піки після диференціювання, логіка предметної області дозволяє визначити сезонну структуру з періодом s=30 та параметрами P=1, Q=0 або 1. З урахуванням цього було сформовано три конфігурації SARIMA-моделей для подальшого порівняння та вибору оптимальної:

- SARIMA(1, 1, 1)(1, 1, 0, 30)
- SARIMA(2, 1, 1)(1, 1, 0, 30)
- SARIMA(1, 1, 1)(1, 1, 1, 30)

За результатами порівняння трьох конфігурацій SARIMA-моделей, найкращі значення інформаційних критеріїв AIC = 22696.84, BIC = 22724.54 показала модель з наступними параметрами SARIMA(1,1,1)(1,1,1,30). Ключові параметри та результати оцінки найкращої моделі наведено на рис. 3.10.

SARIMAX Results

Dep. Variable:	Consum	**No. Observations:**	1947
Model:	SARIMAX(1, 1, 1)x(1, 1, 1, 30)	**Log Likelihood**	-11343.419
Date:	Tue, 06 May 2025	**AIC**	22696.839
Time:	19:29:44	**BIC**	22724.544
Sample:	01-01-2019 - 04-30-2024	**HQIC**	22707.042
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.0145	0.024	0.609	0.542	-0.032	0.061
ma.L1	-0.9223	0.009	-104.301	0.000	-0.940	-0.905
ar.S.L30	-0.0104	0.013	-0.770	0.441	-0.037	0.016
ma.S.L30	-0.9816	0.019	-52.620	0.000	-1.018	-0.945
sigma2	9540.2712	318.743	29.931	0.000	8915.546	1.02e+04

Ljung-Box (L1) (Q):	0.00	**Jarque-Bera (JB):**	1.02
Prob(Q):	0.96	**Prob(JB):**	0.60
Heteroskedasticity (H):	1.15	**Skew:**	-0.03
Prob(H) (two-sided):	0.08	**Kurtosis:**	3.10

Рисунок 3.10 – Розгорнуті результати оцінювання параметрів обраної моделі SARIMA

Усі основні коефіцієнти моделі є статистично значущими ($P < 0.05$), що підтверджує адекватність її структури.

Далі були проаналізовані залишки моделі. Для перевірки відповідності залишків припущенням моделі були побудовані два графіки (рис. 3.11 та рис. 3.12):

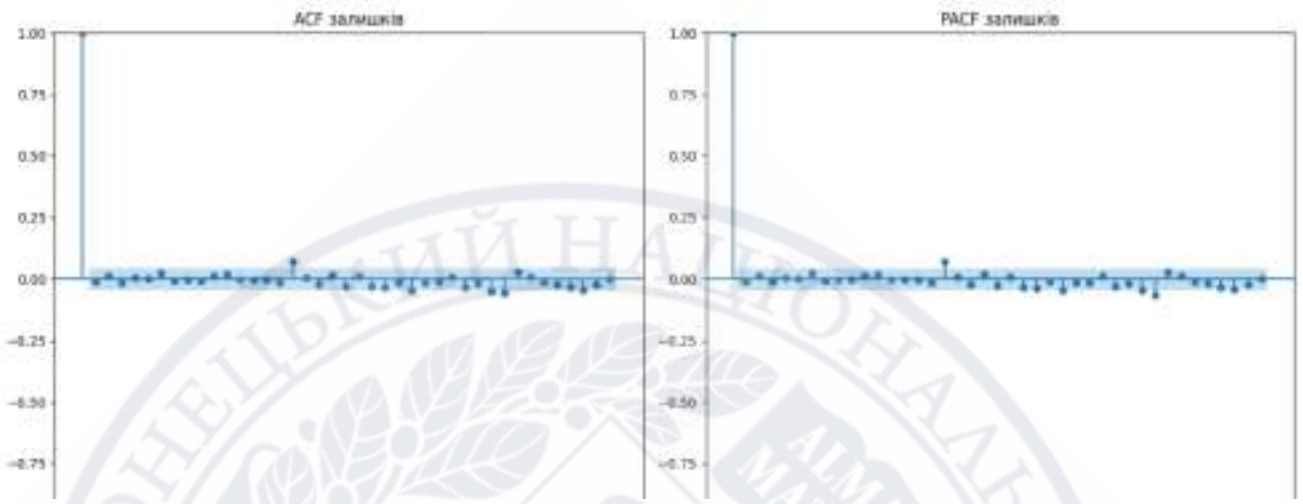

Рисунок 3.11 – Графіки автокореляційної (ACF) та часткової автокореляційної (PACF) функцій залишків

Усі значення перебувають у межах довірчих інтервалів, що свідчить про відсутність автокореляції у залишках.

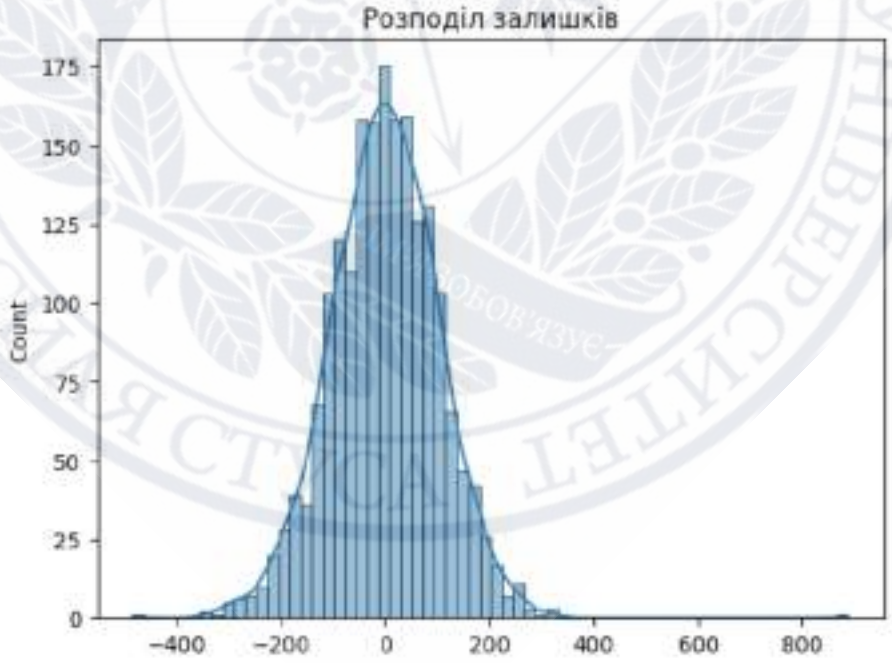

Рисунок 3.12 – Гістограма розподілу залишків SARIMA із накладеною кривою нормального розподілу

Форма розподілу є близькою до нормальної, що підтверджує відповідність припущенням нормальності помилок.

На основі обраної моделі – SARIMA(1,1,1)(1,1,1,30) – було здійснено прогнозування споживання газу на річний тестовий період – з квітня 2024 по березень 2025 року. На графіку (3.13) представлено порівняння прогнозних значень із фактичними даними, а також відображено довірчий інтервал, що ілюструє межі очікуваних коливань. Більшість реальних значень знаходяться всередині довірчого інтервалу, а сам прогноз доволі точно відтворює загальну тенденцію споживання, що вже свідчить про прийнятну якість моделі для середньострокового прогнозування.

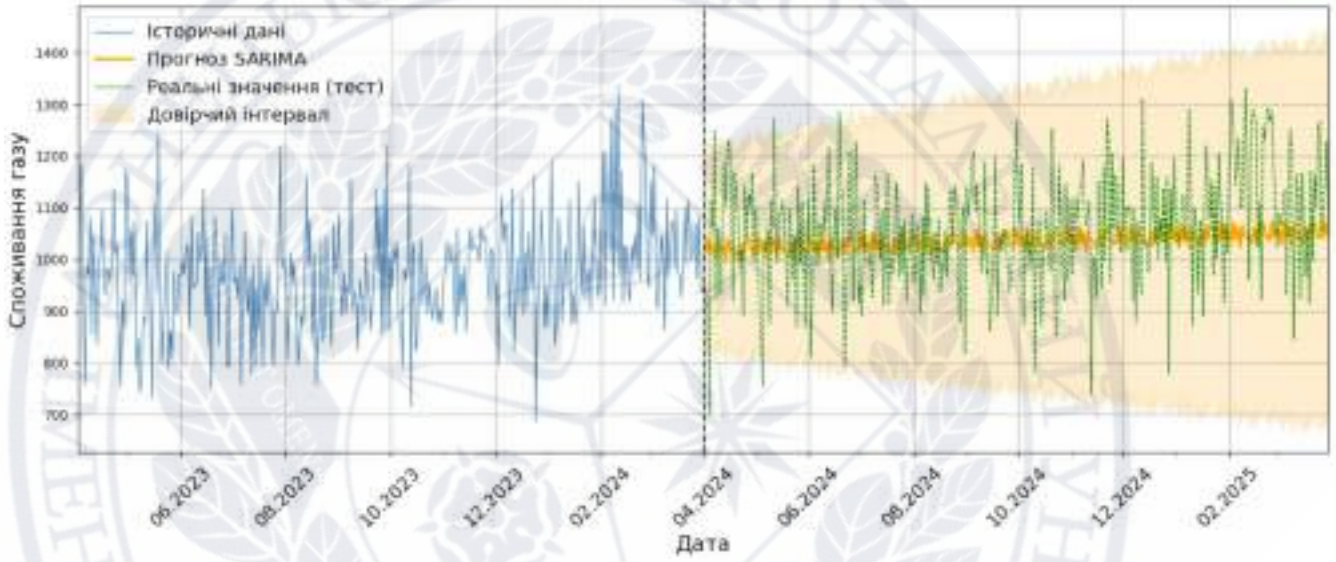

Рисунок 4.13- Прогноз споживання газу на тестових даних моделлю SARIMA

Результати кількісного оцінювання прогнозної якості представлені в таблиці 3.2.

Таблиця 3.2 Результати оцінювання точності прогнозу моделі SARIMA

Метрика	Тренувальні дані	Тестові дані
MAE	81.04	89.78
RMSE	104.05	113.38
sMAPE (%)	8.73	8.53
MASE	0.78	0.86
AIC	22696.84	
BIC	22724.54	

Значення MASE на тестовому наборі (0.86) залишається нижчим за одиницю, що свідчить про вищу точність моделі SARIMA порівняно з наївним прогнозом. Метрики MAE, RMSE та sMAPE демонструють порівняно стабільні значення на тренувальному і тестовому періодах, що вказує на добру узагальнювальну здатність моделі та її придатність для практичного використання.

3.2.3 Налаштування адаптивної нечіткої системи ANFIS

У роботі для налаштування нейро-нечіткої моделі ANFIS використано методи теорії хаосу, які дозволяють визначити оптимальні параметри часової затримки (τ) та розмірності простору станів (m).

Для застосування зазначених методів попередньо було здійснено нормалізацію часового ряду до інтервалу [0; 1] за формулою (2.17). Також для використання програм цього пакету дані конвертуються у формат .dat.

Наступним етапом стала оцінка оптимальної часової затримки (τ) із використанням методу взаємної інформації (mutual information), який дозволяє встановити найінформативніший лаг між спостереженнями (рис. 3.14).

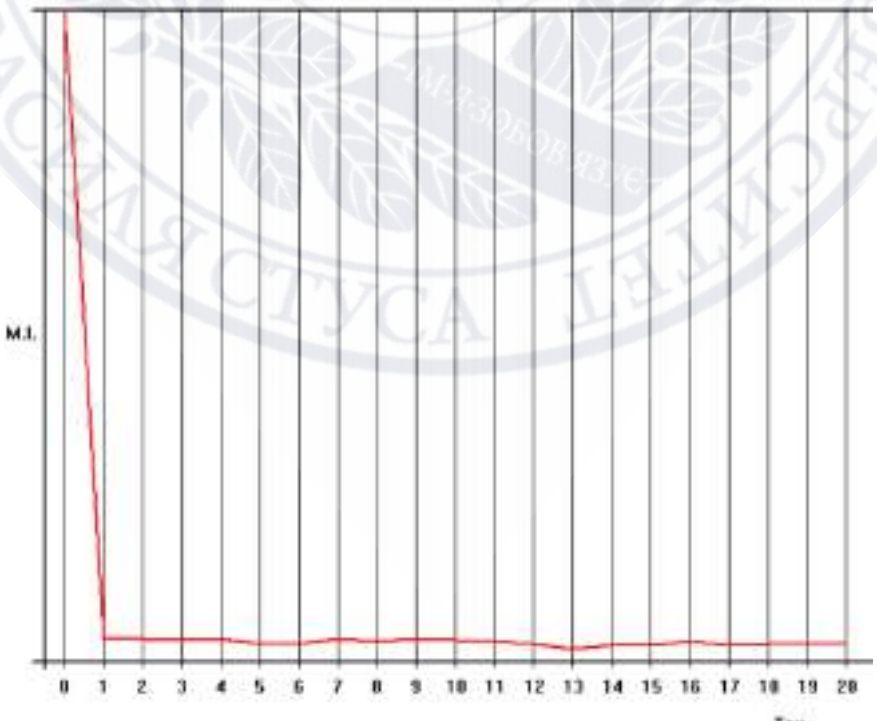

Рисунок 3.14 – Результати виконання програми для пошуку оптимальної часової затримки (mutual)

На основі графіка взаємної інформації видно, що найперше істотне мінімум значення досягається при τ = 1. Це свідчить про те, що для даного часового ряду оптимальною часовою затримкою є 1.

Наступним кроком є визначення оптимальної вбудованої розмірності (m). Для цього було використано метод аналізу хибних найближчих сусідів (FNN), який дозволяє встановити мінімальну розмірність, за якої багатовимірна реконструкція фазового простору стає адекватною (рис. 3.15).

Рисунок 3.15 – Результат виконання програми (fnn) для пошуку оптимальної вбудованої розмірності

Починаючи з m = 5, значення частки практично досягає нуля і залишається стабільно низьким при подальшому збільшенні розмірності. Таким чином, оптимальним вибором для вбудованої розмірності простору є m = 5.

Після визначення оптимальних параметрів часової затримки (τ) та розмірності простору станів (m) було сформовано вибірку навчальних прикладів шляхом побудови векторів, що включають поточне та попередні значення часового ряду у відповідності до обраної затримки.

Для навчання та тестування моделей вибірку було розділено на три підмножини:

- тренувальну, що охоплює дані до квітня 2024 року;

- валідаційну, яка становить 20% від обсягу тренувальної вибірки та використовується для контролю навчання;
- тестову, що охоплює період з квітня 2024 по березень 2025 року та призначена для остаточного оцінювання точності моделі.

Для нейро-нечіткої моделі ANFIS проведено низку експериментів із застосуванням різних типів функцій належності. У кожній моделі для кожної вхідної змінної використовувалися по дві функції належності. Розглядалися три типи функцій:

1. dsigmf (віднімання сигмоїдальних функцій),
2. gaussmf (гаусова функція),
3. gbellmf (генералізована дзвоникоподібна функція).

Навчання моделей здійснювалося на тренувальній вибірці, а якість оцінювалася за валідаційною похибкою. За результатами аналізу найкращі результати показала модель із функціями належності типу dsigmf. Оптимальна кількість епох навчання підбиралася за поведінкою валідаційної похибки для уникнення перенавчання та досягнення найкращої узгодженості моделі з даними.

На рис. 3.16 представлено зміну тренувальної та валідаційної середньоквадратичної помилки (RMSE) в процесі навчання нейро-нечіткої моделі ANFIS.

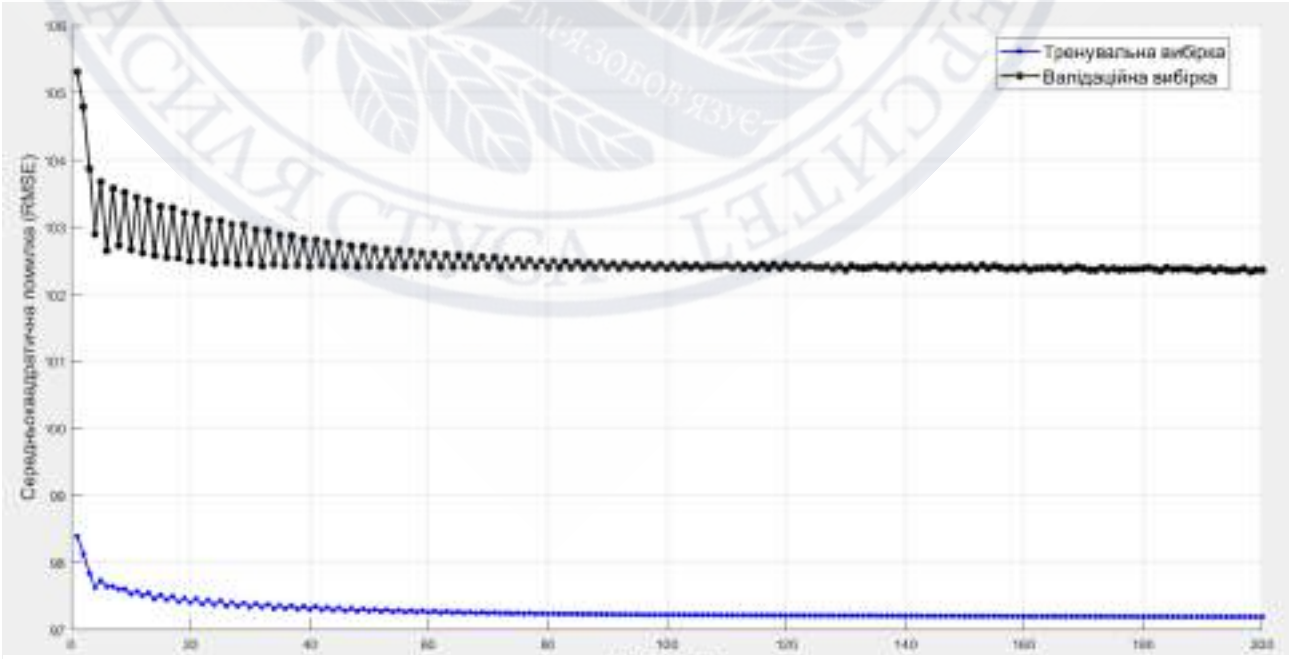

Рисунок 3.16 – Динаміка зміни тренувальної та валідаційної помилок під час навчання ANFIS

Як видно з графіка, тренувальна помилка стабільно зменшується, виходячи на плато після близько 80 епох. Валідаційна помилка також стабілізується, що свідчить про відсутність перенавчання та здатність моделі до адекватного узагальнення даних. Це дозволило обрати оптимальну кількість епох для навчання моделі в 100 епох.

Структуру побудованої нейро-нечіткої моделі представлено на рис. 3.17

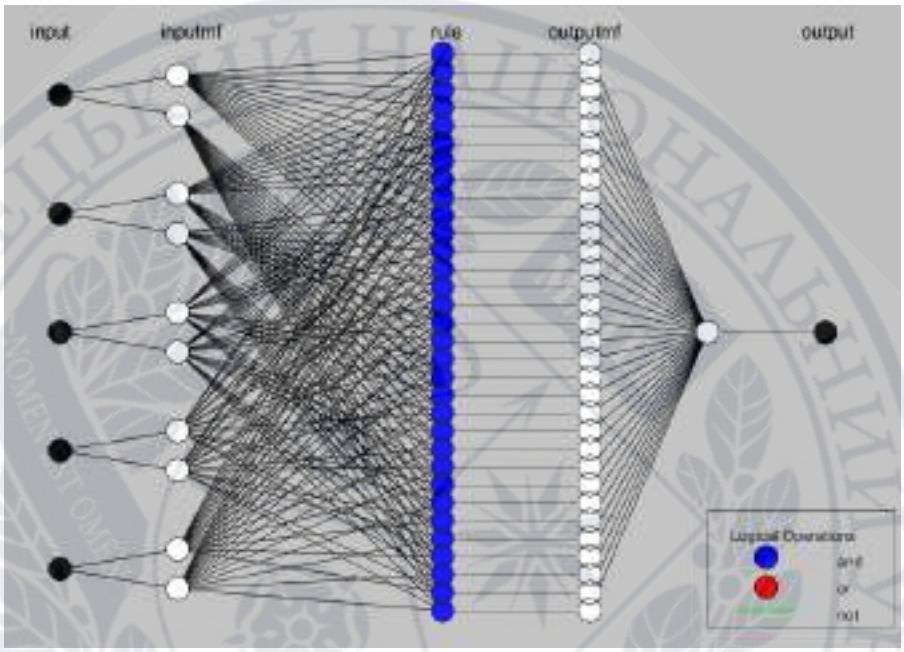

Рисунок 3.17– Структура ANSIS-моделі часового ряду споживання газу

На основі налаштованої нейро-нечіткої моделі ANFIS, побудованої із використанням оптимальних параметрів часової затримки та розмірності простору, було здійснено прогнозування споживання газу на річний тестовий період – з квітня 2024 по березень 2025 року (рис. 3.18).

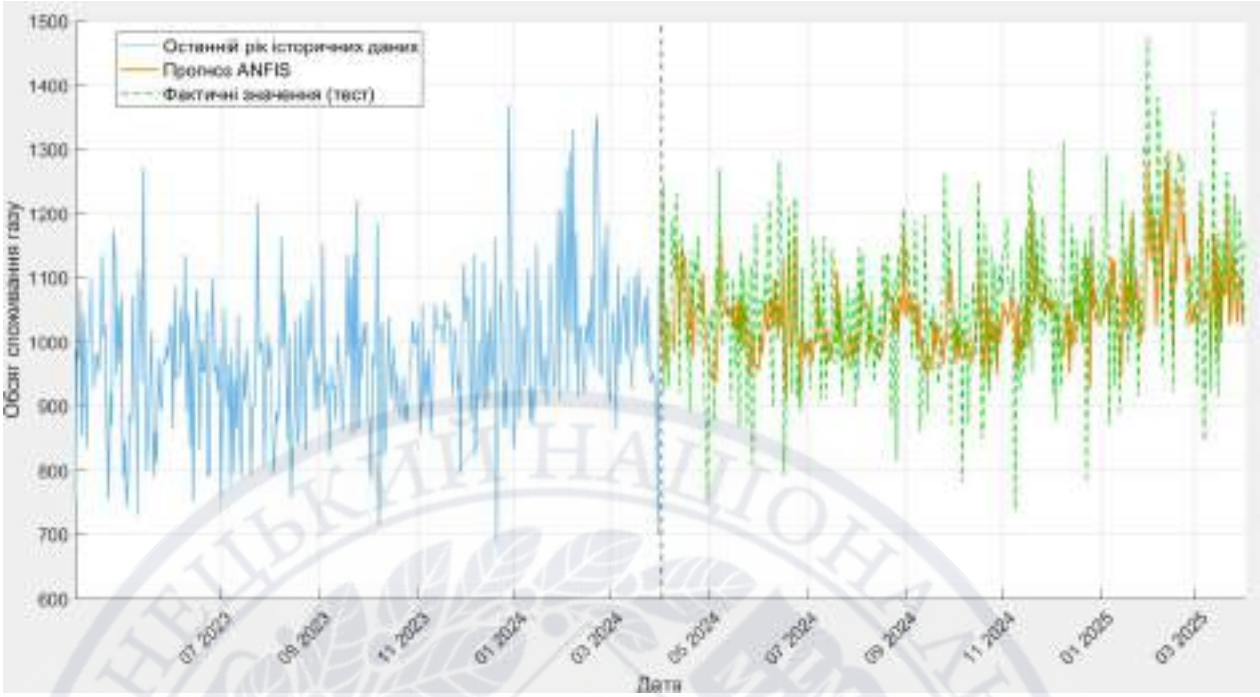

Рисунок 3.18 Прогноз споживання газу на тестових даних моделлю ANFIS

Із графіка видно, що прогнозні значення моделі добре відтворюють основну тенденцію споживання газу протягом усього тестового періоду. Модель загалом слідує за коливаннями реального ряду, хоча локальні пікові та провальні значення не завжди точно збігаються. Прогноз ANFIS демонструє плавний хід, тоді як фактичні дані є більш шумними, що є типовим для реальних часових рядів споживання енергії. Візуально спостерігається адекватна відповідність трендів та сезонних коливань, що свідчить про здатність моделі до узагальнення закономірностей у даних.

Результати кількісного оцінювання прогнозної якості представлені в таблиці 3.3.

Таблиця 3.3 Результати оцінювання точності прогнозу моделі ANFIS

Метрика	Тренувальні дані	Тестові дані
MAE	75.28	79.32
RMSE	97.19	100.23
sMAPE (%)	8.04	7.58
MASE	0.69	0.75
Валідаційні дані	100.35	

Оцінювання точності прогнозу моделі ANFIS показало добрі результати. Значення MAE та RMSE на тестових даних лише трохи вищі за тренувальні, що свідчить про відсутність перенавчання. Низький рівень sMAPE (менше 8 %) підтверджує високу якість прогнозу, а MASE менше одиниці вказує на перевагу моделі над наївним підходом. Помилка на валідаційних даних (100.356) узгоджується з тестовими результатами. Загалом, модель забезпечує надійне прогнозування споживання газу.

3.2.4 Побудова моделі глибокого навчання N-BEATS

Підготовка даних для моделі N-BEATS. Перед побудовою прогнозної моделі N-BEATS було здійснено попередню підготовку даних відповідно до вимог бібліотеки NeuralForecast, яка використовується для реалізації цієї моделі. NeuralForecast передбачає певну структуру вхідного датасету: дані повинні містити три обов'язкові колонки – ds (дата або часовий індекс спостереження), unique_id (ідентифікатор часового ряду) та y (значення спостереження).

Для приведення даних у відповідність до цього формату:

- Було створено новий стовпець unique_id, у якому для всієї вибірки присвоєно значення 'series_1', що позначає одну часову серію.
- Таблиця була реорганізована у формат ['Date', 'unique_id', 'Consum'].
- Після цього проведено перейменування: стовпець Date було перейменовано на ds, а Consum – на y.

Таким чином, датасет було адаптовано для роботи з бібліотекою NeuralForecast, що забезпечує коректне навчання та прогнозування моделлю N-BEATS (рис. 3.19).

	ds	unique_id	y
0	2019-01-01	series_1	887.0
1	2019-01-02	series_1	810.0
2	2019-01-03	series_1	951.0
3	2019-01-04	series_1	612.0
4	2019-01-05	series_1	729.0
...	...	...	...
2277	2025-03-27	series_1	1052.0
2278	2025-03-28	series_1	1109.0
2279	2025-03-29	series_1	1151.0
2280	2025-03-30	series_1	1228.0
2281	2025-03-31	series_1	1067.0

2282 rows × 3 columns

Рисунок 3.19 – Адаптований датасет
для роботи з NeuralForecast

Для побудови прогнозної моделі було обрано горизонт прогнозування тривалістю 30 днів, що відповідає задачам оперативного планування споживання газу та дозволяє ефективно враховувати короткострокові коливання у поведінці часового ряду.

Розмір вхідного вікна (input_size), що визначає кількість попередніх спостережень для генерації одного прогнозу, встановлено як потрійну величину горизонту, тобто HORIZON = 30, INPUT_SIZE = 3 * HORIZON.

Далі для побудови та оцінювання моделі прогнозування N-BEATS дані було розділено на три підмножини: тренувальні дані (до квітня 2024), тестова вибірка (з квітня 2024 до березня 2025) та валідаційна вибірка - для цього було обрано валідаційний розмір тривалістю 365 днів.

Валідаційні дані передавалися через параметри моделі та використовувалися для реалізації механізму ранньої зупинки (early stopping) у разі погіршення валідаційної помилки.

На першому етапі були сформовані три варіанти моделі N-BEATS із різними типами стеків (stack_types):

- модель 1 – лише "seasonality" для врахування сезонних коливань;
- модель 2 – комбінація "seasonality" та "trend" для моделювання сезонності й тренду;
- модель 3 – універсальний стек "generic" для автоматичного налаштування функціональних залежностей.

Кожну модель навчали окремо для оцінки їх здатності відтворювати динаміку споживання газу. Отримані результати стали основою для подальшого вдосконалення архітектури.

На наступному етапі застосовано автоматичний підбір гіперпараметрів за допомогою бібліотеки Optuna. Цей підхід використовувався як допоміжний інструмент для уточнення параметрів моделі, зокрема типу стеків, функції активації, методу нормалізації та кількості змін навчальної ставки. У процесі пошуку було сформовано 20 моделей із різними комбінаціями параметрів, що дозволило оцінити вплив архітектурних налаштувань на точність прогнозування.

Для автоматичного підбору використовувалася спеціально створена конфігурація, частина якої наведена на рис. 3.20.

```
def config(trial):
    return {
        "scaler_type": trial.suggest_categorical("scaler_type", ["minmax", "robust", "standard"]),
        "stack_types": trial.suggest_categorical("stack_types", [["identity"], ["trend"], ["seasonality"],
         ["trend", "seasonality"]]),
        "activation": trial.suggest_categorical("activation", ["ReLU", "LeakyReLU"]),
        "num_lr_decays": trial.suggest_int("num_lr_decays", 0, 3),
        "input_size": trial.suggest_int("input_size", 2 * HORIZON, 3 * HORIZON),
        "random_seed": trial.suggest_int("random_seed",1, 20),
        "batch_size": trial.suggest_categorical("batch_size", [32, 64, 128]),
        "windows_batch_size": trial.suggest_categorical("windows_batch_size", [128, 256, 512]),
        "max_steps": 150,
        "val_check_steps": 10,
        "early_stop_patience_steps": 3,
        "start_padding_enabled": False,
    }
```

Рисунок 3.20 – Конфігурація гіперпараметрів для автоматичного пошуку оптимальних налаштувань моделі N-BEATS за допомогою Optuna

Після створення базових моделей і проведення автоматичного пошуку було зроблено висновки щодо найбільш ефективних поєднань архітектурних рішень для конкретного часового ряду споживання газу.

У результаті аналізу встановлено, що найкращу якість прогнозування забезпечує комбінування трендової та сезонної складових. Тому в якості фінальної конфігурації було обрано наступні налаштування:

- типи стеків (stack_types): ["trend", "seasonality"], що дозволяє моделі одночасно враховувати як довгострокову динаміку споживання газу, так і сезонні коливання;
- функція активації (activation): "LeakyReLU", яка забезпечує кращу обробку спадних ділянок ряду у порівнянні зі стандартною ReLU;
- тип нормалізації даних (scaler_type): "standard", що дозволяє забезпечити стабільне навчання при відсутності екстремальних викидів у даних;
- максимальна кількість кроків тренування (max_steps): 200 із перевіркою валідаційної помилки кожні 10 кроків (val_check_steps=10);
- умова ранньої зупинки навчання (early_stop_patience_steps): 3 кроки без покращення, що дозволяє запобігти перенавчанню моделі.

На основі налаштованої моделі глибокого навчання N-BEATS, було здійснено прогнозування споживання газу на річний тестовий період – з квітня 2024 по березень 2025 року(рисунок 3.21).

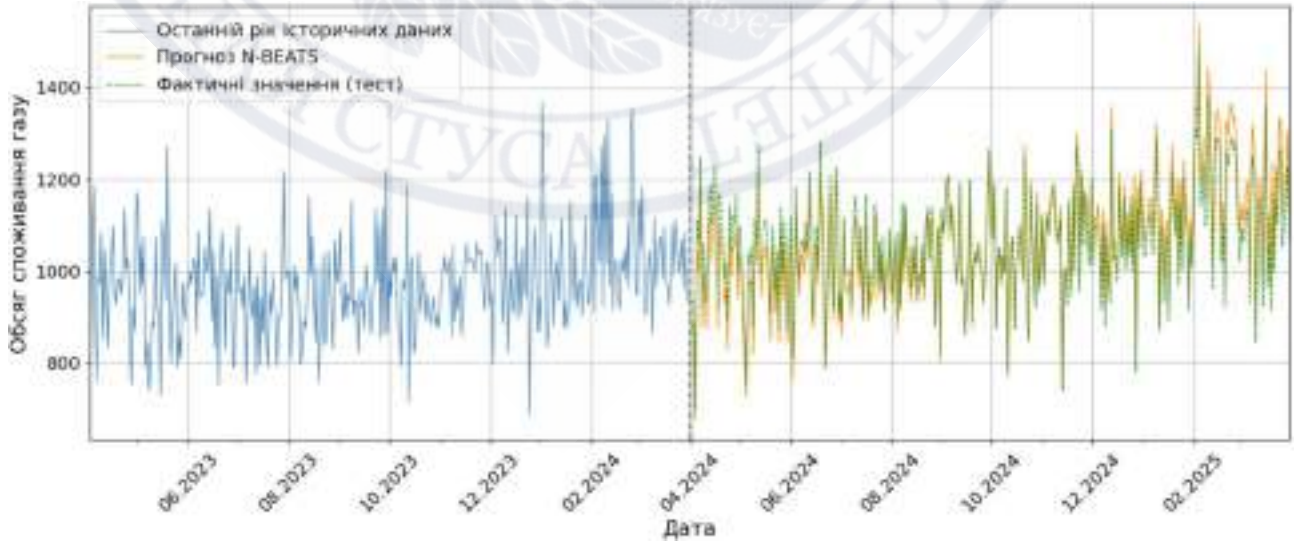

Рисунок 3.21 – Прогноз споживання газу на тестових даних моделлю N-BEATS

Оскільки модель N-BEATS навчалась із горизонтом прогнозування 30 днів, прогнозування річного періоду здійснювалося покроково, шляхом багаторазового розширення часового ряду попередньо прогнозованими значеннями.

На кожному кроці нові спрогнозовані дані доповнювали вихідний часовий ряд, що дозволяло отримати прогноз на тривалу перспективу за допомогою моделі короткострокового прогнозування.

Як видно з графіка, модель досить точно відтворює основні коливання реального споживання газу протягом року, демонструючи хорошу адаптацію до змін сезонності та загальної тенденції розвитку часового ряду.

На графіку залишкової похибки прогнозу моделі N-BEATS (рис. 3.22) видно, що похибки протягом усього тестового періоду залишаються в межах допустимого діапазону. Спостерігається легка тенденція до зменшення похибки впродовж року, що може бути наслідком зміни сезонних коливань у фактичних даних. В цілому залишки рівномірно коливаються навколо нульової лінії без наявності систематичних відхилень, що свідчить про адекватність моделі та відсутність значущого зсуву прогнозу.

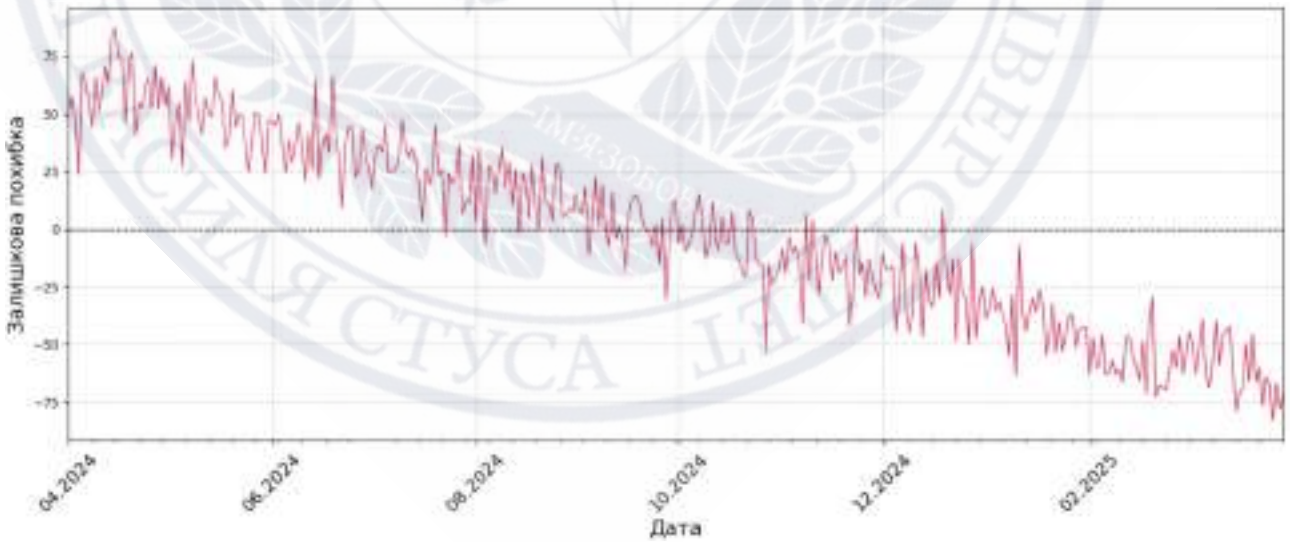

Рисунок 3.22 – Графік залишкової похибки прогнозу моделі N-BEATS на тестових даних

Результати кількісного оцінювання прогнозної якості представлені в таблиці 3.4.

Таблиця 3.4 Результати оцінювання точності прогнозу моделі N-BEATS

Метрика	Тренувальні дані	Тестові дані
MAE	60.08	61.05
RMSE	87.52	90.69
sMAPE (%)	6.04	6.64
MASE	0.53	0.59
Валідаційні дані	90.49	

Результати оцінювання моделі N-BEATS демонструють стабільну якість прогнозу. Значення MAE і RMSE на тестовій вибірці близькі до тренувальних, що свідчить про відсутність перенавчання. Значення sMAPE на тестових даних становить 6.58%, що вказує на високу точність прогнозу. Низьке значення показника MASE також підтверджує ефективність побудованої моделі у порівнянні з наївними методами прогнозування.

3.3 Порівняння точності моделей

З метою вибору найбільш ефективної моделі для прогнозування споживання газу було виконано оцінювання точності чотирьох моделей: Holt-Winters, SARIMA, ANFIS та N-BEATS. Результати порівняння представлені в таблицях 3.5 та 3.6 для тренувальної та тестової вибірок відповідно.

Таблиця 3.5 Оцінювання точності прогнозу моделей на тренувальній вибірці

Моделі	Метрики			
	MAE	RMSE	sMAPE (%)	MASE
Holt-Winters	78.79	98.92	8.42	0.76
SARIMA	81.04	104.05	8.73	0.78
ANFIS	75.28	97.19	8.04	0.69
N-BEATS	60.08	87.52	6.04	0.53

Таблиця 3.6 Оцінювання точності прогнозу моделей на тестовій вибірці

Моделі	Метрики			
	MAE	RMSE	sMAPE (%)	MASE
Holt-Winters	97.69	122.05	9.30	0.90
SARIMA	89.78	113.38	8.53	0.86
ANFIS	79.32	100.23	7.58	0.75
N-BEATS	61.05	90.69	6.64	0.59

Аналіз отриманих результатів показує, що модель N-BEATS демонструє найкращі показники точності як на тренувальній, так і на тестовій вибірках за всіма обраними метриками – MAE, RMSE, sMAPE та MASE.

Серед інших моделей:

- ANFIS показує кращі результати порівняно із Holt-Winters та SARIMA, однак поступається N-BEATS;
- Holt-Winters та SARIMA демонструють дещо вищі значення похибок, що свідчить про менш точне відображення складної структури часового ряду.

Особливо слід відзначити, що різниця між тренувальними та тестовими похибками для моделі N-BEATS є мінімальною, що свідчить про її хорошу узагальнюючу здатність та відсутність перенавчання.

З урахуванням отриманих результатів і практичних аспектів, можна надати такі рекомендації:

- N-BEATS є найбільш доцільним вибором для задач короткострокового та середньострокового прогнозування (горизонт 30 днів і до одного року). Модель демонструє високу точність як при оперативному плануванні, так і при побудові довгострокових прогнозів, що робить її універсальним інструментом. Разом із тим варто враховувати, що N-BEATS вимагає більше обчислювальних ресурсів, налаштування гіперпараметрів та глибшої підготовки даних. Інтерпретованість прогнозу при використанні N-BEATS обмежена, оскільки модель працює як "чорний ящик";

- ANFIS можна рекомендувати у випадках, коли важливо забезпечити кращу інтерпретованість моделі та обмежити обчислювальні витрати. Модель підходить для завдань середньої складності та є компромісом між точністю і пояснюваністю;
- Holt-Winters є найбільш простим у реалізації методом. Його доцільно використовувати для базового прогнозування при стабільних даних або обмеженій кількості спостережень;
- SARIMA придатна для стабільних часових рядів із чітко вираженою сезонністю та трендом, однак вимагає ретельної перевірки стаціонарності та правильного підбору параметрів.

Для демонстрації практичної здатності моделі N-BEATS було побудовано прогноз обсягів споживання газу на один рік вперед (рис. 3.23).

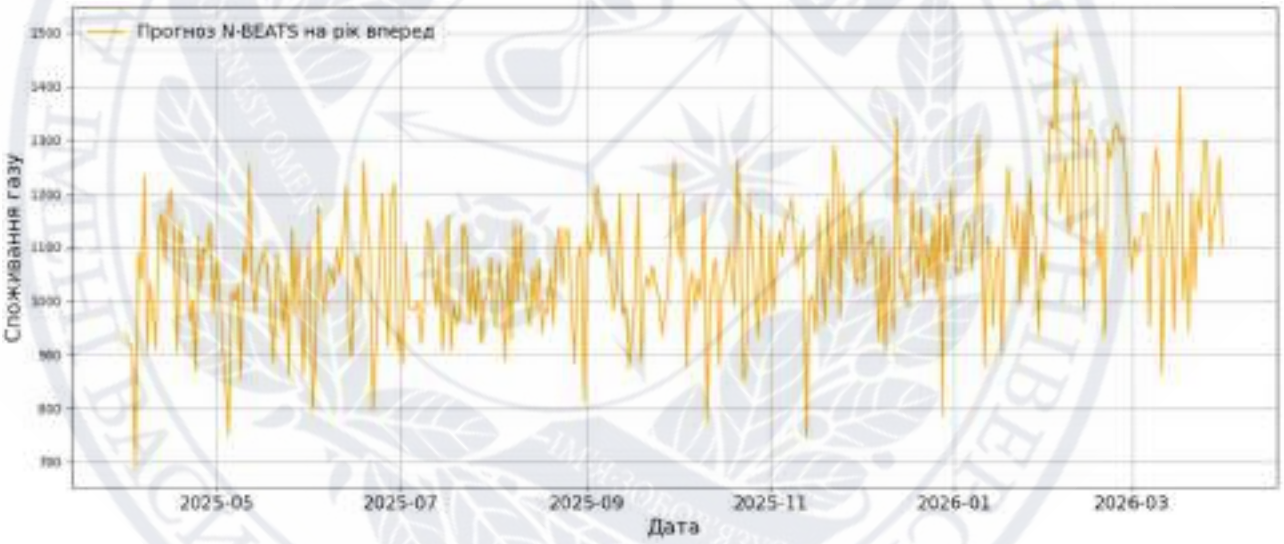

Рисунок 3.23 – Прогноз споживання газу на рік вперед за моделлю N-BEATS

Прогноз відтворює загальний характер сезонних коливань, властивий часовому ряду споживання газу: у літні місяці обсяги споживання зменшуються через зниження потреби в опаленні, а в зимовий період – зростають. Отримані результати підтверджують здатність моделі не лише точно прогнозувати за історичними даними, але й будувати реалістичні сценарії на перспективу, що є важливим для оперативного планування ресурсів підприємства.

ВИСНОВКИ

У дипломній роботі було здійснено комплексне дослідження проблеми прогнозування обсягів споживання природного газу на прикладі реальних даних комунального підприємства. Робота охоплювала повний цикл побудови прогнозної системи: від попередньої обробки даних до розробки, налаштування, тестування та порівняння різних моделей прогнозування часових рядів.

На першому етапі було проведено аналіз та підготовку вхідних даних. Первинні дані представляли собою помісячні обсяги споживання газу за період з 2019 по 2025 рік, що вимірювались у кубічних метрах. Для розширення розмірності навчальної вибірки та забезпечення адекватного функціонування прогнозних моделей дані були перетворені у поденний формат. Таке рішення дозволило поліпшити розпізнавання сезонних та трендових компонентів у часовому ряді, що позитивно вплинуло на якість прогнозування.

У процесі дослідження було побудовано та протестовано чотири моделі прогнозування: модель Хольта-Вінтерса, SARIMA, ANFIS та N-BEATS. Кожна з моделей була ретельно налаштована з урахуванням специфіки вихідних даних та поставленої задачі. Для моделей SARIMA та Holt-Winters було виконано стандартні процедури аналізу стаціонарності та сезонності часового ряду. Модель ANFIS використовувала нечітке моделювання залежностей між вхідними та вихідними даними, а модель N-BEATS базувалася на глибоких нейронних мережах, що дозволило ефективно працювати з необробленими часовими рядами без необхідності ручної побудови ознак.

У процесі побудови моделей особливу увагу було приділено коректному налаштуванню параметрів та відбору конфігурацій, що забезпечують найкращі результати. Для моделі N-BEATS додатково було проведено автоматизований пошук оптимальних гіперпараметрів за допомогою інструмента Optuna, що дозволило підібрати найбільш вдалі комбінації параметрів без суттєвого збільшення часу розробки.

Оцінка якості прогнозування здійснювалася за допомогою поширених метрик: MAE, RMSE, sMAPE та MASE, що дозволяє комплексно оцінити середню абсолютну помилку, середньоквадратичну помилку та відносну помилку моделі. Результати оцінювання на тренувальній та тестовій вибірках показали, що модель N-BEATS забезпечує найкращі показники точності за всіма метриками.

Інші моделі, зокрема Holt-Winters та SARIMA, показали гірші результати, особливо при довгострокових прогнозах, що свідчить про їхню обмежену здатність враховувати складні трендові та сезонні зміни у даних. Модель ANFIS продемонструвала середні результати, однак мала перевагу у простоті інтерпретації отриманих прогнозів.

На підставі отриманих результатів дослідження було розроблено рекомендації щодо використання моделей у практичних задачах прогнозування споживання газу:

- для короткострокового та середньострокового прогнозування (горизонт 30 днів і до одного року) доцільним є використання моделі N-BEATS, яка забезпечує високу точність, адаптивність до складної структури даних та зберігає прийнятну якість прогнозу навіть при збільшенні періоду передбачення;
- для базового прогнозування, яке потребує швидкого налаштування та високої інтерпретованості, можливе використання моделі Holt-Winters;
- у випадках прогнозування стаціонарних або сезонних часових рядів з обмеженим обсягом даних можна розглядати застосування моделі SARIMA;
- модель ANFIS є доцільною для завдань середньої складності, де важливо зберегти баланс між точністю та інтерпретованістю результатів.

Таким чином, за допомогою сучасних методів машинного навчання та аналізу даних, у роботі вдалося досягти високої точності прогнозування, що дозволяє рекомендувати розроблені підходи для практичного застосування у сфері енергетичного планування.

СПИСОК ВИКОРИСТАНИХ ПОСИЛАНЬ

1. А. В. Іваненко, Т. В. Січко. Методи машинного навчання в аналізі та прогнозуванні споживчого попиту. Прикладні аспекти сучасних міждисциплінарних досліджень. 2024. С. 166-168 (дата звернення: 08.04.2025).

2. 2.58. Дані про споживання комунальних ресурсів КП "Центральний міський стадіон" - Datasets - Open data portal. *Головна сторінка - Data.gov.ua*. URL: https://data.gov.ua/en/dataset/2-58-kpcms-resources (дата звернення: 15.02.2025).

3. Про схвалення Стратегії енергетичної безпеки: Розпорядж. Каб. Міністрів України від 4.08.2021 № 907-р. URL: https://zakon.rada.gov.ua/laws/main/907-2021-%25D1%2580#n10 (дата звернення: 08.04.2025).

4. К. В. Гуменюк, Т. В. Січко. Основи машинного навчання. Комп'ютерні технології обробки даних. 2022, м. Вінниця (дата звернення: 29.03.2025).

5. Роман Г. Прогнозування споживання електричної енергії електротехнічних комплексів міської електричної мережі. *International science journal of engineering & agriculture*. 2022. Т. 1, № 3. С. 152–160. URL: https://doi.org/10.46299/j.isjea.20220103.13 (дата звернення: 11.04.2025).

6. Application of neural networks for predicting electric load / V. Kalinchyk et al. *Bulletin of NTU "khpi". series: problems of electrical machines and apparatus perfection. the theory and practice*. 2024. No. 2 (12). P. 50–55. URL: https://doi.org/10.20998/2079-3944.2024.2.10 (дата звернення: 01.04.2025).

7. Athiyarath S., Paul M., Krishnaswamy S. A comparative study and analysis of time series forecasting techniques. *SN computer science*. 2020. Vol. 1, no. 3. URL: https://doi.org/10.1007/s42979-020-00180-5 (дата звернення: 10.04.2025).

8. Camara A., Feixing W., Xiuqin L. Energy consumption forecasting using seasonal ARIMA with artificial neural networks models. *International journal of*

business and management. 2016. Vol. 11, no. 5. P. 231. URL: https://doi.org/10.5539/ijbm.v11n5p231 (дата звернення: 12.04.2025).

9. Chung Y. M., Halim Z. A. The general structure of the Takagi-Sugeno ANFIS model, 2014. *Researchgate*. URL: https://www.researchgate.net/figure/The-general-structure-of-the-Takagi-Sugeno-ANFIS-model_fig1_261476274. (ліцензія https://creativecommons.org/licenses/by-nc-nd/4.0/)

10. Conventional models and artificial intelligence-based models for energy consumption forecasting: a review / N. Wei et al. *Journal of petroleum science and engineering*. 2019. Vol. 181. P. 106187. URL: https://doi.org/10.1016/j.petrol.2019.106187 (дата звернення: 12.04.2025).

11. Deep learning for time series forecasting: advances and open problems / A. Casolaro et al. *Information*. 2023. Vol. 14, no. 11. P. 598. URL: https://doi.org/10.3390/info14110598 (дата звернення: 10.04.2025).

12. De Livera A. M., Hyndman R. J., Snyder R. D. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the american statistical association*. 2011. Vol. 106, no. 496. P. 1513–1527. URL: https://doi.org/10.1198/jasa.2011.tm09771 (дата звернення: 20.04.2025).

13. Dudek G. STD: a seasonal-trend-dispersion decomposition of time series. *IEEE transactions on knowledge and data engineering*. 2023. P. 1–13. URL: https://doi.org/10.1109/tkde.2023.3268125 (дата звернення: 15.04.2025).

14. Enhanced N-BEATS for mid-term electricity demand forecasting / M. Kasprzyk et al. 2024. (Preprint) URL: 10.48550/arXiv.2412.02722.

15. Exponential smoothing model selection for forecasting / B. Billah et al. *International journal of forecasting*. 2006. Vol. 22, no. 2. P. 239–247. URL: https://doi.org/10.1016/j.ijforecast.2005.08.002 (дата звернення: 15.04.2025).

16. Ferbar Tratar L., Strmčnik E. The comparison of Holt–Winters method and Multiple regression method: a case study. *Energy*. 2016. Vol. 109. P. 266–276. URL: https://doi.org/10.1016/j.energy.2016.04.115 (дата звернення: 05.04.2025).

17. Forecasting of demand using ARIMA model / J. Fattah et al. *International journal of engineering business management*. 2018. Vol. 10. P. 184797901880867. URL: https://doi.org/10.1177/1847979018808673 (дата звернення: 11.04.2025).

18. Forecasting: principles and practice. OTexts, 2021. 442 p.

19. Fraser A. M., Swinney H. L. Independent coordinates for strange attractors from mutual information. Physical review A. 1986. Vol. 33, no. 2. P. 1134–1140. URL: https://doi.org/10.1103/physreva.33.1134 (дата звернення: 12.04.2025).

20. Godahewa R., Bergmeir C., I Webb G. A strong baseline for weekly time series forecasting. 2020. (Preprint). URL: https://doi.org/10.48550/arXiv.2010.08158.

21. Hecht M., Zitzmann S. Sample size recommendations for continuous-time models: compensating shorter time series with larger numbers of persons and vice versa. *Structural equation modeling: a multidisciplinary journal*. 2020. P. 1–8. URL: https://doi.org/10.1080/10705511.2020.1779069 (дата звернення: 04.04.2025).

22. Jacob M., Neves C., Vukadinović Greetham D. Short term load forecasting. *Forecasting and assessing risk of individual electricity peaks*. Cham, 2019. P. 15–37. URL: https://doi.org/10.1007/978-3-030-28669-9_2 (дата звернення: 11.04.2025).

23. Jang J. S. R. ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*. 1993. Vol. 23, no. 3. P. 665–685. URL: https://doi.org/10.1109/21.256541 (дата звернення: 16.04.2025).

24. Jaramillo Monge M. Time series analysis for electrical demand forecasting. *Digital technology for smart grid innovative algorithmic solutions for engineering problems*. 2024. URL: https://doi.org/10.17163/abyaups.44.351 (дата звернення: 18.04.2025).

25. Kamolov S., Iskhakov D., Ziyaev B. Machine learning methods in time series forecasting: a review. *Annals of mathematics and computer science*. 2021. Vol. 2. P. 10–14. URL: https://doi.org/10.56947/amcs.v2.13 (дата звернення: 11.05.2025).

26. Karaboga D., Kaya E. Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. *Artificial intelligence review*.

2018. Vol. 52, no. 4. P. 2263–2293. URL: https://doi.org/10.1007/s10462-017-9610-2 (дата звернення: 01.04.2025).

27. Kodba S., Perc M., Marhl M. Detecting chaos from a time series. *European journal of physics*. 2004. Vol. 26, no. 1. P. 205–215. URL: https://doi.org/10.1088/0143-0807/26/1/021 (дата звернення: 13.04.2025).

28. Long-Term forecasting of electrical loads in kuwait using prophet and holt–winters models / A. I. Almazrouee et al. *Applied sciences*. 2020. Vol. 10, no. 16. P. 5627. URL: https://doi.org/10.3390/app10165627 (дата звернення: 16.04.2025).

29. Mamdani and sugeno fuzzy inference systems - MATLAB & simulink. *Mathworks*. URL: https://www.mathworks.com/help/fuzzy/types-of-fuzzy-inference-systems.html (дата звернення: 19.04.2025).

30. Matthew B Kennel, Reggie Brown, Henry DI Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev-1992- A 45- ст. 3403-3411 URL: https://doi.org/10.1103/PhysRevA.45.3403

31. Matjaž perc: time. *Matjaž Perc*. URL: http://www.matjazperc.com/time/ (дата звернення: 10.05.2025).

32. M Competition | Time Series Data - International Institute of Forecasters. *International Institute of Forecasters*. URL: https://forecasters.org/resources/time-series-data/ (дата звернення: 04.04.2025).

33. Modeling and identification of complex systems using neuro-fuzzy techniques.(dept.e) / M. El-Hosiny et al. *MEJ. mansoura engineering journal*. 2021. Vol. 27, no. 1. P. 45–54. URL: https://doi.org/10.21608/bfemu.2021.142599 (дата звернення: 24.04.2025).

34. N-BEATS: neural basis expansion analysis for interpretable time series forecasting / B. N. Oreshkin et al. 2020. (Preprint) URL: https://doi.org/10.48550/arXiv.1905.10437.

35. N-BEATS neural network for mid-term electricity load forecasting / B. N. Oreshkin et al. *Applied energy*. 2021. Vol. 293. P. 116918.

URL: https://doi.org/10.1016/j.apenergy.2021.116918 (дата звернення: 02.04.2025).

36. NBEATS - nixtla. *Nixtlaverse - Nixtla*. URL: https://nixtlaverse.nixtla.io/neuralforecast/models.nbeats.html (дата звернення: 02.04.2025).

37. Peixeiro M. Time series forecasting in python. Manning, 2022. 400 p.

38. Rotshtein A., Pustylnik L., Giat Y. Fuzzy logic and chaos theory in time series forecasting. *International journal of intelligent systems*. 2016. Vol. 31, no. 11. P. 1056–1071. URL: https://doi.org/10.1002/int.21816 (дата звернення: 07.04.2025).

39. Sandhya Arora M. K. Forecasting the future: a comprehensive review of time series prediction techniques. *Journal of electrical systems*. 2024. Vol. 20, no. 2s. P. 575–586. URL: https://doi.org/10.52783/jes.1478 (дата звернення: 10.04.2025).

40. Shiwakoti R. K., Charoenlarpnopparut C., Chapagain K. Time series analysis of electricity demand forecasting using seasonal ARIMA and an exponential smoothing model. *2023 international conference on power and renewable energy engineering (PREE)*, Tokyo, Japan, 20–22 October 2023. URL: https://doi.org/10.1109/pree57903.2023.10370319 (дата звернення: 18.04.2025).

41. Trauth M. H. Time series analysis. *Springer textbooks in earth sciences, geography and environment*. Cham, 2025. P. 189–291. URL: https://doi.org/10.1007/978-3-031-57949-3_5 (дата звернення: 07.04.2025).

42. Usha T. M., Balamurugan S. A. A. Seasonal based electricity demand forecasting using time series analysis. *Circuits and systems*. 2016. Vol. 07, no. 10. P. 3320–3328. URL: https://doi.org/10.4236/cs.2016.710283 (дата звернення: 05.04.2025).

ДОДАТКИ

ДОДАТОК А

Перелік використаних бібліотек

```
# Basic libraries for data processing and visualization
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# For deep learning and neural networks
import torch
import lightning.pytorch as pl

import datetime   # Manipulating data formats

# Classical machine learning and accuracy metrics
from sklearn.linear_model import
LinearRegression
from sklearn.metrics import (mean_absolute_error,
                             mean_squared_error,
                             mean_absolute_percentage_error)

# Time series analysis and statistical model building
import statsmodels.api as sm
from statsmodels.tsa.stattools import adfuller, acf, pacf
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.statespace.sarimax import SARIMAX

# Neural network prediction з NeuralForecast
import
neuralforecast
from neuralforecast.models import NBEATS
from neuralforecast.losses.pytorch import SMAPE

import optuna  # Hyperparameter optimization
```

ДОДАТОК Б

Завантаження даних та візуалізація часового ряду

```
df = pd.read_csv('Test2.csv', parse_dates=['Date'],
index_col='Date')
series = df['Consumption']

# Time series visualization
plt.figure(figsize=(14, 6))
plt.plot(series, color='steelblue')
plt.xlabel("Рік", fontsize=16)
plt.ylabel("Споживання газу", fontsize=16)
plt.grid(True)

plt.xlim(series.index.min(), series.index.max())

plt.xticks(fontsize=14)

plt.tight_layout()
plt.show()
```

ДОДАТОК В

Виявлення та усунення аномалій

```
# Calculating the Z-score
mean = series.mean()
std = series.std()
z_score = (series - mean) / std
threshold = 3

series_smoothed = series.copy()
series_smoothed[np.abs(z_score) > threshold] = np.nan

plt.figure(figsize=(14, 6))
plt.plot(series, label='Часовий ряд')
plt.scatter(anomalies.index, anomalies.values, color='red',
label='Аномалії', zorder=5)
plt.xlabel("Рік", fontsize=14)
plt.ylabel("Споживання", fontsize=14)
plt.legend(fontsize=14)
plt.xticks(fontsize=14)
plt.xlim(series.index.min(), series.index.max())
plt.grid(True)
plt.show()

# Interpolation (smoothing)
series_smoothed = series_smoothed.interpolate(method='linear')


plt.figure(figsize=(14, 6))
plt.plot(series, label='Оригінальний ряд', alpha=0.5)
plt.plot(series_smoothed, label='Ряд зі згладженими аномаліями',
color='green')
plt.xlabel("Рік",fontsize=14)
plt.ylabel("Споживання",fontsize=14)
plt.legend(fontsize=14)
plt.xticks(fontsize=14)
plt.xlim(series.index.min(), series.index.max())
plt.grid(True)
plt.show()


series=series_smoothed
```

ДОДАТОК Д

Декомпозиція часового ряду

```
decomposition = seasonal_decompose(series, model='additive',
period=365)

fig = decomposition.plot()
for ax in fig.axes:
    ax.title.set_fontsize(14)
    ax.yaxis.label.set_fontsize(12)
plt.tight_layout()
plt.show()
```

ДОДАТОК Е

Побудова та оцінка моделі Holt-Winters

```
# Training
model_30_add = ExponentialSmoothing(train, trend='add',
seasonal='add', seasonal_periods=30, damped_trend=True).fit()

residuals_HW = model_30_add.resid

# Ljung-Box test
ljung_test = acorr_ljungbox(residuals_HW.dropna(), lags=[10, 20,
30], return_df=True)
print("Тест Льюнга-Бокса:")
print(ljung_test)

# Histogram of residuals + QQ
sns.histplot(residuals_HW, kde=True)
plt.title("Розподіл залишків")
plt.show()

# sMAPE
def smape(y_true, y_pred):
    return 100 * np.mean(2 * np.abs(y_pred - y_true) /
(np.abs(y_true) + np.abs(y_pred)))

# Forecast for the test period
y_true_test = test
y_pred_test = forecast_hw

# MAE
mae_test = mean_absolute_error(y_true_test, y_pred_test)
# RMSE
rmse_test = np.sqrt(mean_squared_error(y_true_test, y_pred_test))
# sMAPE
smape_test = smape(y_true_test, y_pred_test)
# MASE
mase_test = mae_test / mae_naive
```

ДОДАТОК Ж

Побудова та оцінка моделі SARIMA

```
# ACF, PACF
series1 = series.diff().dropna()

fig, axes = plt.subplots(2, 1, figsize=(12, 8))

plot_acf(series1, ax=axes[0], lags=50)
axes[0].set_title('ACF (Автокореляційна функція)')

plot_pacf(series1, ax=axes[1], lags=50, method='ywm')
axes[1].set_title('PACF (Часткова автокореляційна функція)')

plt.tight_layout()
plt.show()

# Training
model = SARIMAX(train,
                order=(1, 1, 1),
                seasonal_order=(1, 1, 1, 30),
                enforce_stationarity=False,
                enforce_invertibility=False)

res_SARIMA = model.fit()
res_SARIMA.summary()

# Residue diagnostics
residuals = res_SARIMA.resid
plt.figure(figsize=(14, 6))
plt.subplot(1, 2, 1)
plot_acf(residuals.dropna(), lags=40, ax=plt.gca(), title='ACF
залишків')
plt.subplot(1, 2, 2)
plot_pacf(residuals.dropna(), lags=40, ax=plt.gca(), title='PACF
залишків')
plt.tight_layout()
plt.show()

# Histogram of residuals + QQ
sns.histplot(residuals, kde=True)
plt.title("Розподіл залишків")
plt.show()

# Ljung-Box test
ljung_test = acorr_ljungbox(residuals.dropna(), lags=[10, 20, 30],
return_df=True)
print("Тест Льюнга-Бокса:")
print(ljung_test)
```

ДОДАТОК К

Побудова та оцінка моделі ANFIS

```
% Configuring FIS generation parameters
opt_gen = genfisOptions("GridPartition");
opt_gen.NumMembershipFunctions = 2;  % Кількість МФ на вхід
opt_gen.InputMembershipFunctionType = "dsigmf";  % Тип МФ (можна
'trimf', 'sigmf' тощо)

% Creating an initial FIS system
initFIS = genfis(trainInput, trainOutput, opt_gen);

% Setting training parameters
opt = anfisOptions;
opt.InitialFIS = initFIS;
opt.EpochNumber = 100;
opt.ValidationData = valData;

opt.DisplayANFISInformation = 1;
opt.DisplayErrorValues = 1;
opt.DisplayStepSize = 1;
opt.DisplayFinalResults = 1;

% Training
[fisout, trainError, stepSize, chkFIS, valError] =
anfis([trainInput trainOutput], opt);

% Forecast for the test period
y_true_test = testOutput;
y_pred_test = evalfis(fis, testInput);

% MAE
mae_test = mean(abs(y_true_test - y_pred_test));
% RMSE
rmse_test = sqrt(mean((y_true_test - y_pred_test).^2));
% sMAPE
smape_test = 100 * mean(2 * abs(y_pred_test - y_true_test) ./ ...
                (abs(y_pred_test) + abs(y_true_test)));
% MASE
mase_test = mae_test / mase_denominator;
```

ДОДАТОК Л

Побудова та оцінка моделі N-BEATS

```
# Horizon and input window
HORIZON = 30
INPUT_SIZE = 3 * HORIZON

# Model
model_trend_seasonality = NBEATS(
    h=HORIZON,
    input_size=INPUT_SIZE,
    stack_types=["trend", "seasonality"],
    activation="LeakyReLU",
    scaler_type="standard",
    num_lr_decays=0,
    max_steps=200,
    val_check_steps=10,
    early_stop_patience_steps=3,
    start_padding_enabled=False,
)

gas_forecast = fcst.predict()

# Forecast for the test period
y_true = test['y'].values
y_pred = merged['NBEATS'].values

# MAE
mae_nbeats = np.mean(np.abs(y_true - y_pred))
# RMSE
rmse_nbeats = np.sqrt(np.mean((y_true - y_pred) ** 2))
# sMAPE
smape_nbeats = 100 * np.mean(2 * np.abs(y_true - y_pred) /
(np.abs(y_true) + np.abs(y_pred)))
# MASE
train_values = train['y'].values
mase_denominator = np.mean(np.abs(train_values[1:] -
train_values[:-1]))
mase_nbeats = mae_nbeats / mase_denominator
```