

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»

Кафедра інформатики факультету інформатики

Прогнозування енергоспоживання будівель

за допомогою машинного навчання

Текстова частина до курсової роботи
за спеціальністю 122 «Комп'ютерні науки»

Керівник курсової роботи

Ст.в. Салата К. В.

(прізвище та ініціали)

(підпис)

“ ____ ” _____ 2025 р.

Виконав студент КН-3

Телегін Н. А. “ ____ ” _____ 2025 р.

м. Київ – 2025 рік

Календарний план

Тема: Прогнозування енергоспоживання будівель за допомогою машинного навчання

Календарний план виконання роботи:

№ п/п	Назва етапу курсової роботи	Термін виконання етапу	Примітка
1	Отримання завдання на курсову роботу.	05.11.2024	
2	Огляд технічної літератури за темою роботи.	11.02.2025	
3	Аналіз проблематики прогнозування часових рядів	23.02.2025	
4	Аналіз сучасних статистичних методів прогнозування	13.03.2025	
5	Аналіз сучасних методів прогнозування з використанням штучного інтелекту	19.03.2025	
6	Аналіз сучасних ансамблевих методів	23.04.2025	
7	Програмування моделей.	04.04.2025	
8	Аналіз отриманих результатів.	20.04.2025	
9	Написання пояснювальної роботи.	30.04.2025	
10	Створення слайдів для доповіді та написання доповіді.	01.05.2025	
10	Аналіз отриманих результатів з керівником	03.05.2025	
11	Корегування роботи	04.05.2025	

Студент Телегін Н.А. _____

Керівник Салата К.В. _____

“ ___ ” _____ 2025 р.

ЗМІСТ

Вступ	5
Розділ 1. Аналіз існуючих методів (алгоритмів) вирішення поставленої задачі	8
Розділ 2. Прогнозування енергоспоживання будівель як часовий ряд.	12
2.1. Теоретичні засади прогнозування часових рядів	12
2.1.1 Основні компоненти часових рядів.	12
2.1.2. Прогнозування енергоспоживання на денному часовому ряді.	14
2.1.3. Прогнозування енергоспоживання на тижневому часовому ряді.	16
2.1.4. Прогнозування енергоспоживання на річному часовому ряді.	18
2.2. Опис обраних моделей прогнозування енергоспоживання будівель.	19
2.2.1. Прогнозування енергоспоживання з використанням статистичних моделей.	19
2.2.2. Прогнозування класичними методами машинного навчання.	22
2.2.3. Прогнозування ансамблевими методами на основі дерев рішень.	23
2.2.4. Прогнозування енергоспоживання з використанням нейронних мереж.	26
2.3 Проблематика статичних моделей прогнозування енергоспоживання будівель.	28
2.4. Вхідні дані та аналіз отриманих результатів.	31
2.4.1. Обрана вибірка даних.	31
2.4.2. Результати статистичних моделей.	31
2.4.3. Результати моделей з використанням машинного навчання.	32
Розділ 3. Експериментальна частина.	34
3.1 Будова створеної гібридної моделі.	34
3.2 Отримані результати.	37
Висновки	40
Посилання та використана література	41

Перелік використаних термінів та умовних позначень.

LSTM - довготривала короткочасна пам'ять (Long Short-Term Memory)

DNN - глибока нейронна мережа (Deep Neural Network)

ANN - штучна нейронна мережа (Artificial Neural Network)

PSO - оптимізація рою частинок (Particle Swarm Optimization)

ICA - імперіалістичний конкурентний алгоритм (Imperialist Competitive Algorithm)

TLBO - оптимізація на основі викладання й навчання (Teaching-Learning-Based Optimization)

ACO - оптимізація колонії мурашок (Ant Colony Optimization)

XGB - екстремальне посилення градієнта (eXtreme Gradient Boosting)

DT - дерево рішень (Decision Tree)

RF - випадковий ліс (Random Forest)

SVR - регресія опорного вектора (Support Vector Regression)

KNN - K-найближчих сусідів (K-Nearest Neighbors)

ШНМ – штучна нейронна машина

PSO – оптимізація рою частинок (Particle swarm optimization)

Вступ

Прогнозування енергоспоживання будівель – це актуальна задача сучасності, що пов'язана з поняттями енергетичної ефективності, економії ресурсів та екології. Сьогодні прогнозування енергоспоживання є повноцінним полем для дослідження, адже, враховуючи особливості споживання електроенергії та важливість підтримки сталих потужностей та напруги в мережі для її роботи, передбачення витрат у кожний дискретний момент дає змогу ефективніше використовувати та розподілювати енергоресурси.

Конкретизуючи, в умовах глобальних кліматичних змін та нестабільної енергетичної ситуації, особливо в Україні, точне прогнозування енергоспоживання дає змогу оптимізувати використання електроенергії, зменшувати навантаження на енергомережу та уникати перевантажень, а отже в перспективі уникнути аварійних відключень та придумати ефективні стабілізаційні механізми.

Важко заперечити той факт, що сучасні міста стають дедалі енергоємнішими, а будівлі споживають значну частку електроенергії, що вимагає ефективних інструментів для аналізу їхньої енергетичної поведінки. Згідно з даними наукового дослідницького центру Precedence Research [1], ринок енергоприладів з 2024 року до 2034 року зросте в 2 рази, склавши 1346,07 мільярда доларів США. Вже зараз проводяться постійні дослідження [2] та спроби ефективного прогнозування споживання будівель ще на моменті будівництва, адже кожна будівля впливає на загальний енергетичний профіль міста.

У контексті стрімкого розвитку цифрового середовища електронні системи керування створюють підґрунтя для впровадження розумних механізмів прогнозу. Вже зараз в Україні реалізуються проекти, що значно скорочують витрати електроенергії. Так, згідно з даними Економічної правди [2], після встановлення розумних теплотічильників в одному з районів Києва вдалося зменшити витрати за один опалюваний сезон на 23%. Залучення ж

технологій машинного навчання до енергетичного сектору уможливило розробку індивідуалізованих схем споживання як для окремих житлових одиниць, так і для багатоквартирних комплексів, що суттєво розширює потенціал розумного зменшення енергоспоживання.

При цьому розгляд цієї задачі з прикладного наукового погляду має міцне підґрунтя, адже прогнозування енергоспоживання має потенціал через сезонний характер, який пов'язаний з особливостями ведення людської діяльності. Згідно з даними Eurostat [3], у 2022 році домогосподарства в Європі становили 25,8% кінцевого споживання енергії континенту, а основою їхнього енергоспоживання є опалення будинків, що складало 63,5% від кінцевого споживання енергії в житловому секторі. При цьому електроенергія, що використовується для освітлення та більшості електроприладів, становила 13,9%, а частка, яка використовується для нагріву води, була дещо вищою і становила 14,9%. Основні пристрої для приготування їжі вимагали 6,3% енергії, тоді як охолодження приміщень та інші кінцеві споживання охоплювали 0,6% та 0,9% відповідно. Варто одразу зазначити, що опалення будинків має чітку річну сезонність, у той час як електроенергія, що використовується для освітлення, електроприладів, готування їжі, нагріву води – денну сезонність. Саме тому предметною областю цього дослідження є аналіз взаємозалежностей всередині споживчих потреб та спосіб поєднання внутрішніх закономірностей для спроби розв'язання задачі регресії.

Робота складається з трьох розділів.

Перший розділ присвячено аналізу існуючих методів та алгоритмів вирішення задачі прогнозування енергоспоживання будівель. Розглянуто теоретичні засади підходів до прогнозування та математичний апарат моделювання енергоспоживання.

В другому розділі детально досліджується прогнозування енергоспоживання будівель як часовий ряд. Розглянуто основні компоненти часових рядів та особливості прогнозування на різних часових горизонтах: денному, тижневому та річному. Наведено опис різних моделей прогнозування:

статистичних, класичних методів машинного навчання, ансамблевих методів на основі дерев рішень та моделей з використанням нейронних мереж. Проаналізовано проблематику статичних моделей та вхідні дані, а також представлено результати застосування різних методів прогнозування.

Третій розділ присвячено експериментальній частині дослідження. Описано будову створеної гібридної моделі, проаналізовано отримані результати та розглянуто виклики прогнозування енергоспоживання будівель в українському контексті. Виконано порівняльний аналіз запропонованої гібридної моделі з іншими методами прогнозування.

Науковою новизною у цій роботі варто визначити постановку задачі, як задачі прогнозування енергоспоживання на годинному часовому ряді з обмеженням у використанні часових лагів при машинному навчанні. Оскільки при розв'язанні часової регресії в контексті енергоспоживання ми не можемо мати уявлення про попередні години i , найчастіше, вимушені спиратися на дані попередніх місяців та років, було поставлено задачу навчитися прогнозувати на наступний місяць без використання годинних, денних, тижневих лагів, спиратися винятково на історичне споживання. У таких умовах було проаналізовано різні методи вирішення поставленої задачі з використанням різних моделей прогнозування: статистичних, штучних нейронних та моделей на основі дерев рішень.

Розділ 1. Аналіз існуючих методів (алгоритмів) вирішення поставленої задачі

Дослідницьке поле прогнозування енергоспоживання впродовж останніх десятиліть збагатилося різноманітними підходами до моделювання. Сучасний методологічний арсенал структурно розподіляється на три фундаментальні категорії: класичні статистичні методики, нейронні мережеві моделі та ансамблеві моделі.

Науковці в роботі [4] запропонували методологію прогностичного моделювання електроспоживання житлових об'єктів із застосуванням глибинних навчальних алгоритмів. Імплементована модель LSTM продемонструвала надзвичайно високий показник точності (97%) порівняно з DNN та звичайною ANN, завдячуючи комплексному опрацюванню історичних, кліматичних та архітектурно-будівельних параметрів. Попередня підготовка масивів даних мінімізувала похибки обчислень, а верифікація на додаткових вибірках для різнотипних будівель підтвердила узагальнювальну спроможність запропонованого підходу. Матеріали цього дослідження становлять цінну аналітичну базу, однак автор акцентує увагу на залежності прогностичної точності від повноти й якості вхідних даних, а також регіональних кліматичних особливостей.

У дослідженні [5] представлено інтегровану метаевристичну архітектуру штучної нейронної мережі для точнісного прогнозування енергетичних витрат житлових комплексів на початкових етапах проектування, що спрямована на спрощення обчислювальних процесів при збереженні прогностичної надійності. Для підвищення ефективності ШНМ впроваджено оптимізаційні алгоритми рою частинок (PSO), імперіалістичного конкурентного механізму (ICA), навчально-викладацької оптимізації (TLBO) та колонії мурашок (ACO). Дослідницький фокус охопив 16 критичних змінних, включаючи геометричні параметри споруд, просторову орієнтацію, конструктивні характеристики зовнішніх елементів та температурні режими. Експериментальна апробація

шести конфігурацій вхідних параметрів виявила домінантний вплив просторових вимірів будівлі та коефіцієнта сонячного теплонадходження на прогностичну точність. Порівняльний аналіз засвідчив перевагу моделі ANN-TLBO, яка демонструє найвищу точність із нормалізованою середньоквадратичною похибкою 0,033 кВт·год та коефіцієнтом детермінації 93,6%. Варто зауважити, що досягнуті показники потребували значного обсягу даних та багатопараметричного аналізу, що не завжди операційно доступно. Водночас, це дослідження підкреслює кореляцію між якістю початкового збору даних, їх аналітичним опрацюванням та результуючою прогностичною точністю.

У дослідженні [6] представлено інноваційну ансамблевую модель прогнозування енергоспоживання, що інтегрує екстремальне посилення градієнта (XGB), дерево рішень (DT) і випадковий ліс (RF) у стековій архітектурі. Гіперпараметричну оптимізацію всіх складових моделей здійснено за допомогою байєсівського підходу, що максимізує їхню прогностичну ефективність. Проведений автором комплексний кількісно-якісний аналіз підтвердив домінування запропонованої методики над звичними підходами, включаючи регресію опорних векторів (SVR), метод К-найближчих сусідів (KNN), штучні нейромережеві структури (ANN) та інші регресійні алгоритми, демонструючи конкурентоспроможність порівняно з існуючими науковими напрацюваннями. Автор акцентує увагу на тому, що досягнення високої прогностичної точності стало можливим завдяки імплементації комбінаторного підходу, який синтезує різнотипні моделі машинного навчання. Такий методологічний симбіоз призводить до ефективнішого враховування впливів на енергетичний профіль будівель і підвищує достовірність прогнозних оцінок.

Варто наголосити, що значну частку загального енергоспоживання становить опалення та кондиціонування приміщень, що обґрунтовує необхідність відокремленого дослідження методологій прогнозування цього енергетичного сегмента. У дослідженні [7] проаналізовано ефективність двох

поширених прогностичних методик для оцінки енергоспоживання кондиціонерами, що розширило розуміння поведінкових патернів мешканців та методологічних обмежень існуючих підходів. Встановлено, що протягом періоду охолодження температурний режим приміщень підтримувався в діапазоні 25–29°C у 90% випадків активного кондиціонування, з середньою тривалістю функціонування 3,99 години щодня, що істотно перевищує показник опалювального сезону — 1,63 години. Частотність використання кондиціонування повітря в охолоджувальний період була в 5,4 раза вищою порівняно з опалювальним. Енергоспоживчий аналіз на основі польових вимірювань виявив суттєву варіативність користувацьких звичок, з середніми показниками 2,62 кВт·год/день у період охолодження та 0,47 кВт·год/день під час опалення. Натомість метод декомпозиції "зверху вниз" продемонстрував значне завищення показників, оцінюючи середнє споживання на рівні 5,66 кВт·год/день та 4,81 кВт·год/день відповідно. Моделювання методом "знизу вгору" також характеризувалося завищенням результатів, перевищуючи фактичні дані у 2,76 рази через спрощені припущення щодо режимів експлуатації, тоді як метод "зверху вниз" завищив прогноз у 2,89 рази через включення енергоспоживання інших електроприладів. Отримані результати засвідчують істотні прогностичні відхилення обох методологій, що підкреслює критичний вплив поведінкових факторів на точність енергетичних оцінок. Відтак, при моделюванні на рівні індивідуальних користувачів першочергового значення набуває детермінація частотно-інтенсивнісних характеристик використання обігрівального обладнання.

Дослідження [8] презентує аналітичний огляд електроспоживання латвійських домогосподарств, де встановлено, що середньомісячний показник у 2022 році сягав 160 кВт·год, з яких 9% припадало на побутову електротехніку. Дослідницька вибірка охопила 31 житлову одиницю протягом трирічного періоду з урахуванням демографічного чинника. Визначено диференційовані енергетичні профілі: однокімнатне помешкання споживає 480 кВт·год щорічно, двокімнатне — 970 кВт·год, трикімнатне — 1480 кВт·год, а

середньостатистичне індивідуальне споживання становить 547 кВт·год. Ці емпіричні дані, враховуючи географічну та кліматичну схожість з українським контекстом, становлять цінний аналітичний матеріал для поточного дослідження, одночасно акцентуючи на важливості ефективного енергоспоживання не лише як оптимізаційного заходу, але й як платформи для імплементації прогресивних технологічних рішень.

У наступному дослідженні [9] запропоновано модель прогнозування енергоспоживання будівель на основі аналізу часових рядів і опорної векторної машини (SVM). Метою цього дослідження було покращення точності оцінки споживання відновлюваної енергії. Результати експериментальної перевірки продемонстрували високий рівень точності моделі, з втратою 1,5% у навчальному наборі та 4,1% у тестовому, а також максимальну точність 95,5% у тесті на точність нейронної мережі. Аналіз похибки показав, що модель має найменшу відносну похибку прогнозування – 2,3%, тоді як для традиційної SVM вона становила 6,8%, а для моделей хаотичних часових рядів – 4,1%. Отримані результати свідчать про значне покращення здатності прогнозування порівняно з існуючими методами, що робить запропоновану модель перспективною для практичного застосування у сфері прогнозування споживання відновлюваної енергії в будівлях. Однак автор зазначає, що хоча модель й може передбачити споживання енергії в реальних наборах даних, вона не була перевірена за допомогою даних у реальному часі. У той же час аномалії даних, викликані аномальним попитом на споживання, не враховувалися в процесі розробки моделі, а отже вона може не бути настільки прогресивною та точною в наших теперішніх умовах.

Розділ 2. Прогнозування енергоспоживання будівель як часовий ряд.

Часовий ряд – це ряд спостережень y_1, y_2, \dots, y_n випадкової величини (t), утворених у послідовні моменти часу t_1, t_2, \dots, t_n . [10]

2.1. Теоретичні засади прогнозування часових рядів

Аналіз та передбачення споживання енергії будівлями відноситься до складних завдань прогнозування часових рядів, що вимагає детального вивчення історичних даних для виявлення закономірностей та формування обґрунтованих прогнозів. У сфері енергоспоживання будівель часовий ряд формується з набору історичних вимірювань електроспоживання, що перебувають під впливом багатьох екзогенних чинників: температурного режиму навколишнього середовища, часу доби, дня тижня, сезонності та раптових подій (аварійних ситуацій, планових чи позапланових відключень).

2.1.1 Основні компоненти часових рядів.

Енергоспоживання як часовий ряд визначається чотирма ключовими компонентами:

Тренд – довготривала динаміка зміни рівня споживання енергії. Для прикладу, у міських агломераціях можна спостерігати поступове зростання енергоспоживання через збільшення чисельності населення, розбудову інфраструктурних об'єктів та поширення електричного транспорту. Натомість, впровадження енергоощадних технологій та альтернативних джерел енергії може призводити до зниження загального споживання у певних регіонах [12]

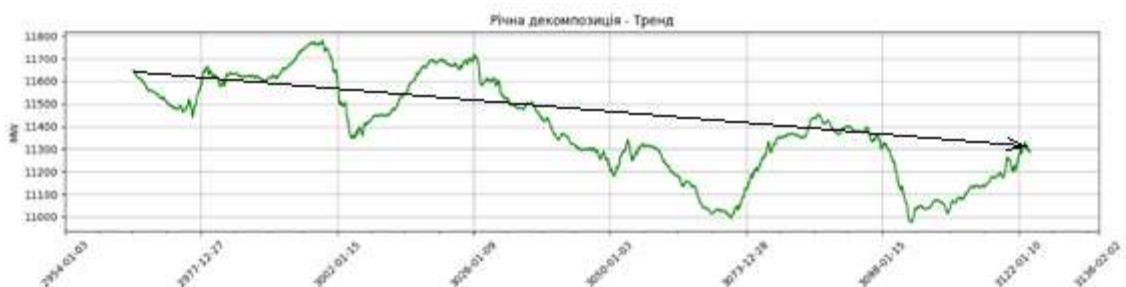


Рис. 1. Тренд енергоспоживання в м. Дейтон, США

Сезонність – регулярні коливання енергоспоживання, що повторюються через визначені часові інтервали. Говорячи про сезонність, варто зауважити, що у зимовий період електроспоживання зростає через масове використання опалювальних приладів, а в літній – через експлуатацію систем охолодження. Так у дослідженні [11] вплив опалення та охолодження в загальному споживанні енергії визначалася такою, що коливається від 18% до 73%, тобто хоч і в різній степені, але, безумовно, високо. Сезонність також чітко простежується в добовому розрізі: пікові навантаження припадають на денні години активності організацій та підприємств, тоді як нічний період характеризується суттєвим зниженням споживання.

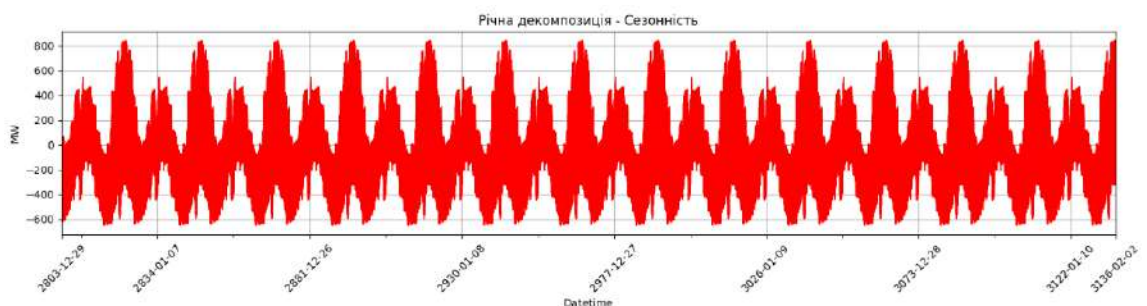


Рис. 2. Сезонність енергоспоживання в м. Дейтон, США.

Циклічність – закономірні відхилення в енергоспоживанні, що мають повторюваний характер, проте без чіткої періодичності. Ілюстрацією можуть слугувати економічні рецесії, що спричиняють зниження енергоспоживання через згорання виробничих потужностей. В умовах військового конфлікту в Україні циклічність набуває особливих форм: спостерігаються різкі падіння споживання під час цілеспрямованих атак на об'єкти енергетичної інфраструктури з подальшим поступовим відновленням після проведення ремонтно-відновлювальних робіт.

Шум – стохастичні, нерегулярні коливання енергоспоживання, що не піддаються чіткому прогнозуванню. Іноді викиди можуть бути настільки високими, що маскують тенденцію та сезонність. [13] Такі відхилення можуть бути зумовлені технічними аваріями, раптовими метеорологічними змінами, непередбаченими навантаженнями на електромережу чи іншими факторами випадкового характеру. Ключовим завданням у підвищенні точності прогнозів

є ефективне відокремлення шумових компонентів від структурованих закономірностей у масиві даних.

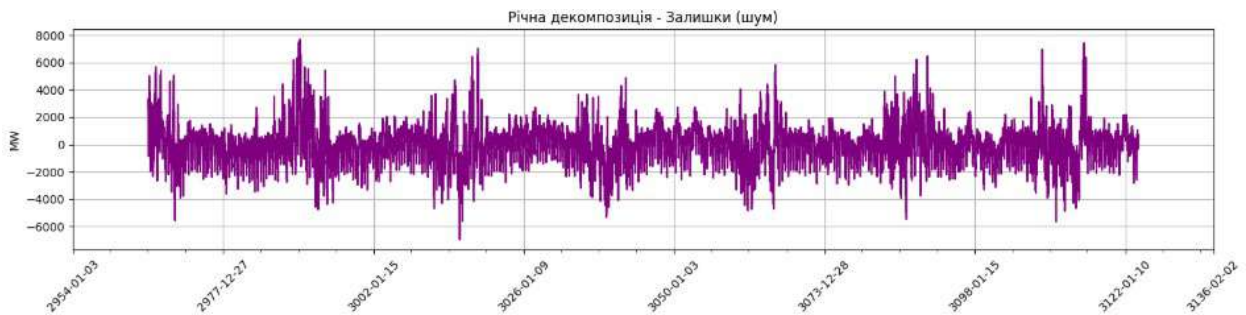


Рис. 3. Шум на графіку енергоспоживання м. Дейтон, США

При цьому прогнозування енергоспоживання потребує також деяких вимог до даних, особливої ваги в цьому контексті набуває поняття стаціонарності – властивості, що віддзеркалює сталість статистичних параметрів ряду незалежно від часової точки спостереження. [10]

Для перевірки цієї властивості енергетичних часових рядів застосовується низка специфічних статистичних інструментів, розширений тест Дікі-Фуллера, що перевіряє наявність одиничних коренів та KPSS-тест, який оцінює сталість середнього значення та дисперсії ряду

Коли аналіз виявляє нестаціонарну природу енергоспоживання, варто скористатися математичними перетвореннями даних – диференціювання послідовних значень або логарифмічної трансформації, щоб привести ряд до форми, придатної для коректного моделювання.

2.1.2. Прогнозування енергоспоживання на денному часовому ряді.

Задачу прогнозування енергоспоживання будівель можна розглядати на різних рівнях часових рядів, такі дані мають чітко виражену сезонність і циклічність, яка проявляється на різних рівнях — добовому, тижневому та сезонному.

Добовий профіль споживання електроенергії істотно варіюється залежно від типу забудови — житлової, промислової чи комерційної. Кожен з цих типів має характерні особливості, зумовлені специфікою діяльності користувачів у межах відповідних об'єктів.

У житловій забудові (житлові будинки, квартири) добовий графік енергоспоживання має двопікову структуру. Перший пік споживання припадає на ранкові години, коли мешканці прокидаються, користуються освітленням, побутовими електроприладами, здійснюють приготування їжі тощо. Після цього, у період з 8:00 до 16:00–17:00, споживання значно знижується, що пов'язано з відсутністю мешканців у помешканнях протягом робочого дня. Другий, зазвичай вищий, пік припадає на вечірній період (приблизно 19:00–21:00), коли відбувається повернення мешканців додому та активне використання енергоспоживчих приладів. У нічний час (після 23:00) споживання електроенергії знижується до мінімального рівня.

У промисловій забудові структура енергоспоживання, як правило, зосереджена в межах стандартного робочого дня (з 7:00 до 18:00). Зростання споживання починається вже з 6:00 у зв'язку із запуском обладнання та підготовкою виробничих процесів. Високий рівень споживання зберігається протягом усього робочого дня, а після 18:00 різко знижується. У нічний час споживання, зазвичай, незначне або відсутнє, якщо підприємство не функціонує у змінному режимі. У випадку безперервного (цілодобового) виробництва добовий графік має більш рівномірний характер, з незначним спадом у нічні години.

Для комерційної забудови (офісні приміщення, заклади торгівлі, торговельно-розважальні центри, адміністративні будівлі) характерним є однопіковий профіль споживання. Активне енергоспоживання починається в межах 8:00–9:00, з відкриттям офісів і торговельних закладів, та досягає максимуму в обідній або післяобідній період (приблизно 11:00–14:00). Після 17:00–18:00 спостерігається поступове зниження споживання, а в нічний час воно залишається на низькому рівні. У вихідні дні профіль споживання може змінюватися: для торговельних об'єктів характерні денні піки вищої інтенсивності, тоді як офісна нерухомість демонструє значне зниження споживання.



Рис. 4. Середнє енергоспоживання по годинах доби м. Дейтон, США.

2.1.3. Прогнозування енергоспоживання на тижневому часовому ряді.

На відміну від добових профілів, тижневий графік енергоспоживання має більш складну структуру, що зумовлюється не лише типом забудови, а й соціально-економічними, поведінковими та організаційними чинниками. Основним джерелом варіацій у тижневому профілі є відмінності між робочими та вихідними днями, режим роботи об'єктів, сезонні особливості та навіть культурно-побутові звички споживачів.

У житловій забудові тижневий профіль енергоспоживання характеризується відносно стабільним загальним рівнем протягом тижня, однак з помітними відмінностями між буднями та вихідними. У будні дні графік відображає типовий ритм життя працюючого населення — ранкові та вечірні піки активності. У вихідні, особливо в суботу та неділю, спостерігається зсув піків: ранковий починається пізніше, а вечірній може бути більш розтягнутим у часі. Також у вихідні дні можливе загальне підвищення денного споживання через триваліше перебування мешканців удома, активніше використання електроприладів, приготування їжі, розважальні заходи тощо. Водночас у деяких випадках загальне енергоспоживання може знижуватися, якщо частина мешканців відїжджає на відпочинок або до заміських будинків.

У промисловій забудові тижневий графік є більш регламентованим і залежить від організації виробничого процесу. Підприємства з п'ятиденним робочим тижнем демонструють стійке високе споживання в робочі дні (понеділок–п'ятниця) та різке зниження або повну відсутність споживання у вихідні (субота–неділя). У випадку підприємств із змінним або безперервним графіком роботи (наприклад, металургія, харчова промисловість, енергетика) тижневий профіль є більш рівномірним, із незначними коливаннями залежно від інтенсивності виробництва, графіку змін, технічного обслуговування або ремонтів. Окремі пікові навантаження можуть також виникати у зв'язку з виконанням планових виробничих завдань на початку або наприкінці тижня.

У комерційній забудові тижневий профіль відзначається значною варіативністю, що залежить передусім від функціонального призначення об'єкта. Для офісних приміщень характерне енергоспоживання у робочі дні з чітко вираженим зниженням у вихідні, коли приміщення зазвичай не експлуатуються. Натомість об'єкти роздрібної торгівлі, особливо торговельно-розважальні центри, можуть демонструвати підвищене споживання у вихідні, зокрема в суботу та неділю, коли фіксується найбільший потік відвідувачів. У такому разі пікові навантаження можуть бути вищими, ніж у будні. Крім того, окремі види комерційної нерухомості (наприклад, об'єкти сфери послуг, що працюють цілодобово) мають майже рівномірний тижневий профіль із незначною варіацією.

На тижневу структуру енергоспоживання також впливають сезонні чинники (опалювальний сезон або періоди з підвищеним використанням кондиціонерів), святкові дні, які можуть викликати як тимчасове зниження (зупинка виробництва, зачиненість офісів), так і підвищення споживання (святкова активність у житловому секторі чи збільшення потоку покупців у комерційних об'єктах). Погодні умови, такі як різке похолодання або спека, також можуть суттєво змінити звичні тижневі профілі, особливо в енергоємних сферах.

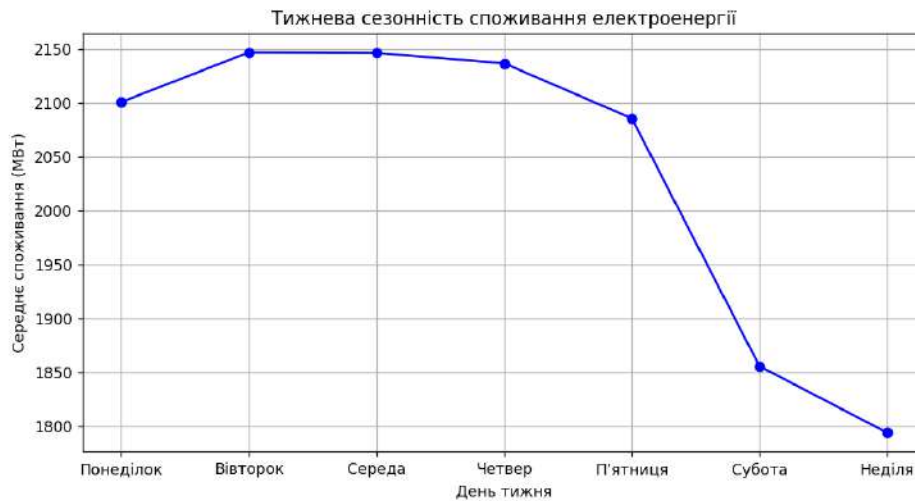


Рис. 5. Середнє енергоспоживання по днях м. Дейтон, США.

2.1.4. Прогнозування енергоспоживання на річному часовому ряді.

Додатково до цього додається річна сезонність, пов'язана з опалювальним сезоном узимку або охолодженням улітку. В контексті України ми можемо спостерігати чітке зростання енергоспоживання в ці періоди, проте сам графік при деталізації може істотно різнитися через вплив кліматичних чинників та їхньої нестабільності – пікові значення можуть припадати на різні тижні та навіть на різні місяці щороку.



Рис 6. Середнє енергоспоживання по місяцях м. Дейтон, США.

При цьому, температура повітря — один з головних зовнішніх чинників, який сильно впливає на споживання: чим сильніше відхилення від комфортної температури, тим більше енергії витрачається на підтримання мікроклімату. Безумовно, також вплив можуть мати вологість, швидкість вітру, хмарність, які опосередковано впливають на опалення або кондиціонування.

2.2. Опис обраних моделей прогнозування енергоспоживання будівель.

2.2.1. Прогнозування енергоспоживання з використанням статистичних моделей.

Говорячи про прогнозування з використанням статистичних моделей, варто звернути увагу на дві основні вимоги до даних для успішного прогнозування – це властивості стаціонарності часового ряду та автокореляція. Якщо стаціонарність була пояснена раніше, то друга властивість потребує окремого опису. Явище автокореляції [13] розкриває внутрішню структуру зв'язків між показниками енергоспоживання, зафіксованими у різні моменти. Ця характеристика дає змогу виявити неочевидні закономірності у споживанні енергоресурсів, зокрема, відмінності між робочими та вихідними періодами використання будівлі.

Математичні методи дослідження автокореляційних зв'язків складаються з таких основних підходів, як розрахунок функції автокореляції (ACF), що відображає залежність між поточними та попередніми значеннями та побудову функції часткової автокореляції (PACF), яка виключає вплив проміжних спостережень.

Ці інструменти відіграють вирішальну роль у визначенні оптимальних параметрів прогностичних моделей сімейства ARIMA, зокрема, через PACF встановлюється порядок авторегресійної складової, що формує основу математичної моделі енергоспоживання.

ARIMA (Autoregressive Integrated Moving Average)

ARIMA – це один із найпоширеніших статистичних методів аналізу та прогнозування часових рядів. Він базується на комбінації авторегресії (AR), інтегрованого процесу (I) для усунення трендів та ковзного середнього (MA).

Моделі позначаються як ARIMA (p,d,q), де 'p' представляє порядок авторегресивної частини, 'd' позначає ступінь першого диференціювання і 'q' позначає порядок ковзної середньої. Авторегресивна частина моделі з порядком 'p' може бути представлена як:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

Модель ковзної середньої порядку 'q':

$$Y_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-p} + \varepsilon_t, \text{ де}$$

Y - це вихідні дані часового ряду

ε_t - ряд похибок. [13]

Даний метод ефективний для моделювання стаціонарних часових рядів і добре підходить для випадків, коли енергоспоживання будується на стабільних довгострокових тенденціях. Його перевага – інтерпретованість і здатність моделювати автокореляцію в даних, що важливо при аналізі періодичних змін у споживанні. Проте в контексті нашої роботи (рис. 1), метод ARIMA не підходить через денну та річну сезонність даних, тому побудований цією моделлю прогноз швидко зводиться до середнього значення (Regression toward the mean). Інакші результати модель ARIMA здатна продемонструвати при апроксимації до місячного часового датафрейму.

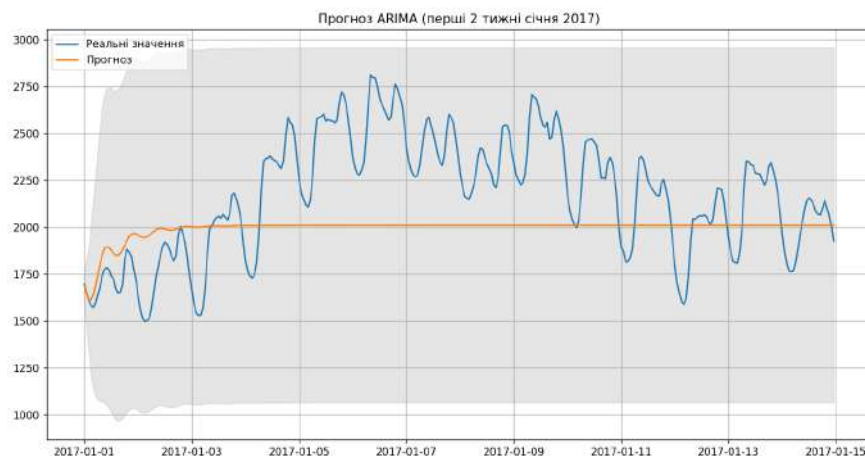


Рис 7. Результати роботи методу ARIMA

SARIMA (Seasonal AutoRegressive Integrated Moving Average) — це розширення класичної моделі ARIMA, яке дає змогу враховувати сезонність у часовому ряді. На відміну від ARIMA, яка базується лише на внутрішній структурі самого ряду, SARIMA надає змогу моделювати більш складні часові процеси. Базова ARIMA не враховує повторювані сезонні структури, що унеможливорює її ефективне застосування в умовах, коли ряди демонструють виражені добові, тижневі або річні цикли, як це часто буває з енергоспоживанням. SARIMA розширює цю модель за рахунок сезонних

компонентів (S), де додається окрема трійка параметрів (P, D, Q) і вказується період сезонності (наприклад, 24 для годинного ряду з добовою періодикою). Така властивість дає змогу моделі коректно враховувати циклічність і, в порівнянні з ARIMA, SARIMA, має більшу гнучкість і складність.

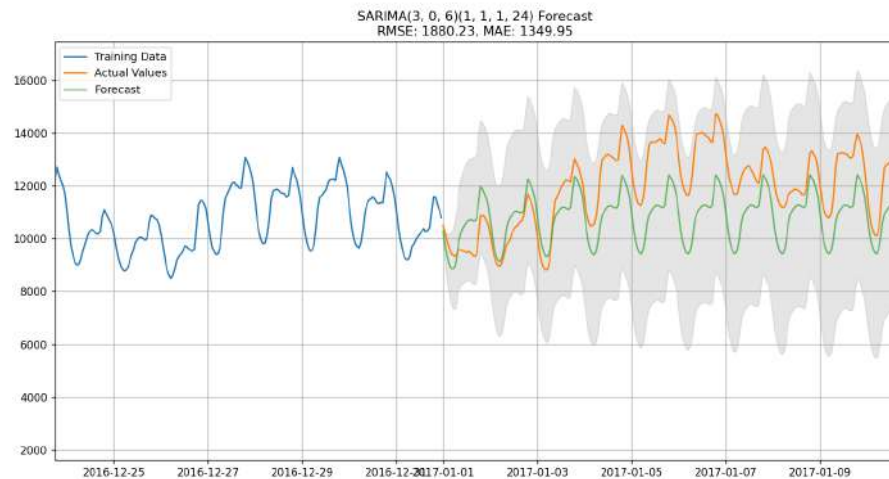


Рис 8. Результати роботи методу SARIMA.

Можна побачити, що SARIMA вловив денний патерн, енергоспоживання, але досі не здатний дискретно розглядати часовий ряд. Видозмінюючи прогноз в конкретних випадках, він лише формує загальний шаблон, але не уточнює його. Тому останнім покращенням в межах статистичних методів варто вважати рахування екзогенних змінних.

Модель SARIMAX.

SARIMAX — це розширення моделі SARIMA, яке здатне враховувати не лише автокореляції і сезонність у часовому ряді, а й зовнішні фактори (екзогенні змінні), які можуть впливати на прогнозоване значення. Тобто SARIMA працює лише з самим рядом і його сезонною структурою, тоді як SARIMAX дає змогу додатково враховувати, наприклад, температуру, день тижня або інші зовнішні ознаки, що можуть бути важливими для точності прогнозу.

У цьому наборі даних, що був використаний для навчання SARIMAX моделі, як екзогенні змінні використано метеорологічні показники — температура, вологість, атмосферний тиск, напрямок і швидкість вітру в Чикаго, а також календарні ознаки — година доби, день тижня та святковість.

Всі ці змінні не є частиною самого ряду енергоспоживання, але можуть впливати на нього, тому використовуються для підвищення точності прогнозування в моделях типу SARIMAX або інших, що враховують зовнішні чинники.

2.2.2. Прогнозування класичними методами машинного навчання.

Лінійна регресія — це статистичний метод моделювання взаємозв'язку між залежною змінною та однією або кількома незалежними змінними. Цей метод належить до базових інструментів прогнозування та аналізу даних, який широко застосовується у різних галузях науки, економіки та інженерії завдяки своїй простоті та інтерпретованості.

У випадку простої лінійної регресії математична модель має вигляд:

$$y = \beta_0 + \beta_1 x + \varepsilon, \text{ де:}$$

y — залежна змінна

x — незалежна змінна

β_0 — вільний член

β_1 — коефіцієнт нахилу

ε — випадкова похибка моделі

Для випадку кількох незалежних змінних формула розширюється до:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon, \text{ де:}$$

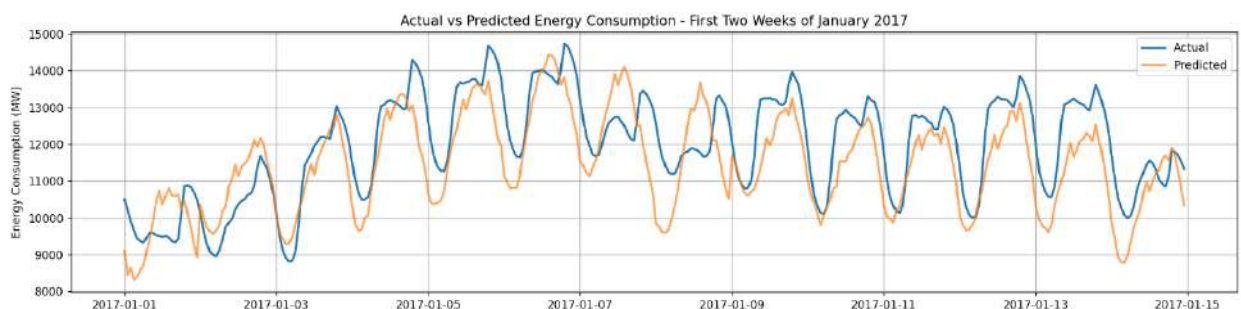
y — залежна змінна

x_1, x_2, \dots, x_n — незалежні змінні

$\beta_0, \beta_1, \beta_2, \dots, \beta_n$ — параметри моделі

ε — випадкова похибка

В контексті нашої задачі, лінійна регресія є базовим методом для прогнозування, що встановлює лінійні залежності між незалежними змінними



(факторами, що впливають на споживання енергії) та залежною змінною (самим споживанням). Вона допомагає виявити основні тенденції впливу площі будівлі, середньої температури та робочих днів на загальне енергоспоживання.

Рис. 9. Результати роботи Лінійної регресії.

2.2.3. Прогнозування ансамблевими методами на основі дерев рішень.

Ансамблеві методи є потужним підходом у машинному навчанні, який полягає в комбінуванні кількох моделей для отримання кращого прогнозу, ніж може забезпечити кожна окрема модель. В контексті прогнозування часових рядів енергоспоживання будівель, ансамблеві методи на основі дерев рішень демонструють виняткову ефективність завдяки здатності вловлювати складні нелінійні залежності, властиві енергетичним даним.

Random Forest (випадковий ліс)

Random Forest представляє собою ансамбль дерев рішень, побудованих із застосуванням двох ключових принципів: bagging (bootstrap aggregating) та випадкового відбору ознак.

Математична модель прогнозування за допомогою Random Forest може бути представлена як:

$$\hat{y} = (1/T) \cdot \sum f_{b(x)}, \text{ де:}$$

\hat{y} — прогнозоване значення

T — кількість дерев в ансамблі

$f_{b(x)}$ — прогноз t-го дерева для вхідних даних x

Особливості Random Forest для прогнозування енергоспоживання:

1. Висока стійкість до перенавчання завдяки усередненню прогнозів багатьох дерев
2. Здатність обробляти велику кількість факторів впливу на енергоспоживання
3. Вбудована оцінка важливості змінних.

4. Ефективна робота з нелінійними залежностями, характерними для енергетичних даних

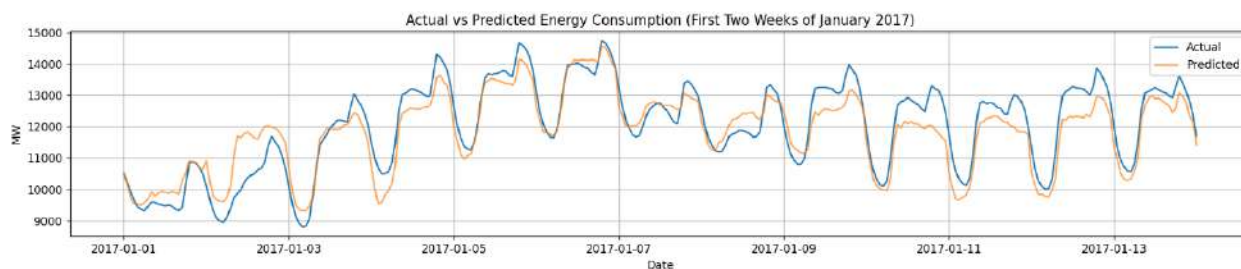


Рис. 10. Результати роботи Random Forest моделі.

Gradient Boosting Machines (GBM)

Gradient Boosting є послідовним ансамблевим методом, який будує модель покроково, кожного разу фокусуючись на помилках попередніх ітерацій.

Математичний запис алгоритму GBM:

$$F_0(x) = \arg \min_{\gamma} \sum L(y_i, \gamma)$$

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x), \text{ де:}$$

$F_m(x)$ — модель на m -й ітерації

$h_m(x)$ — слабкий предиктор (дерево рішень), який мінімізує функцію втрат

η — швидкість навчання

L — функція втрат

Переваги GBM для прогнозування енергоспоживання: послідовне вдосконалення моделі; можливість використання різних функцій втрат, адаптованих під специфіку задачі; висока ефективність при виявленні сезонних компонент енергоспоживання.

XGBoost (Extreme Gradient Boosting)

XGBoost є покращеною версією градієнтного бустингу дерев рішень, що забезпечує високу точність прогнозування та ефективність роботи навіть на великих вибірках. Його ключовою особливістю є здатність зменшувати перенавчання завдяки регуляції та обробці викидів у даних. У випадку прогнозування енергоспоживання XGBoost корисний для роботи з великим

набором вхідних змінних, включаючи погодні фактори, особливості будівель та календарні особливості.

Цільова функція XGBoost:

$$Obj(\theta) = \sum l(y_i, \hat{y}_i) + \sum \Omega(f_k), \text{ де:}$$

l — функція втрат

Ω — регуляризаційний член

f_k — функція k -го дерева

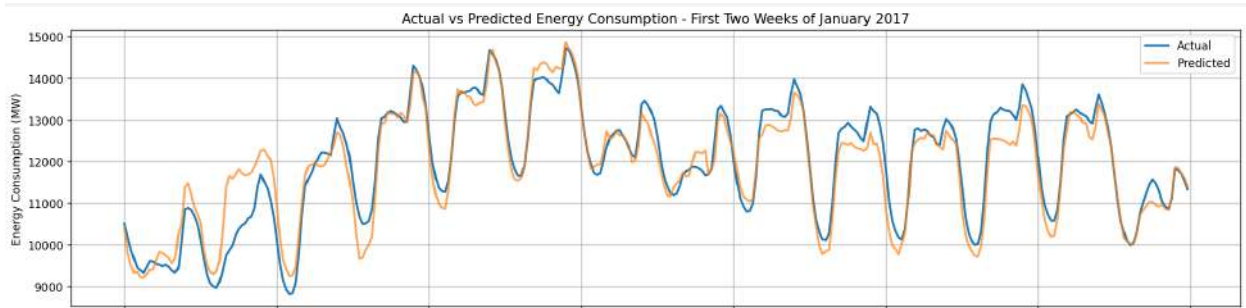


Рис. 11. Результати роботи XGBoost моделі.

LightGBM

LightGBM — це ефективна реалізація градієнтного бустингу, що використовує алгоритм побудови дерев на основі гістограм та стратегію росту дерев за листками.

Ключові особливості LightGBM у контексті прогнозування енергоспоживання: алгоритм GOSS (Gradient-based One-Side Sampling) для ефективного відбору інформативних спостережень; техніка EFB (Exclusive Feature Bundling) для роботи з високорозмірними даними; висока швидкість тренування при збереженні точності прогнозування; зменшене використання пам'яті, що важливо для довготривалих часових рядів енергоспоживання

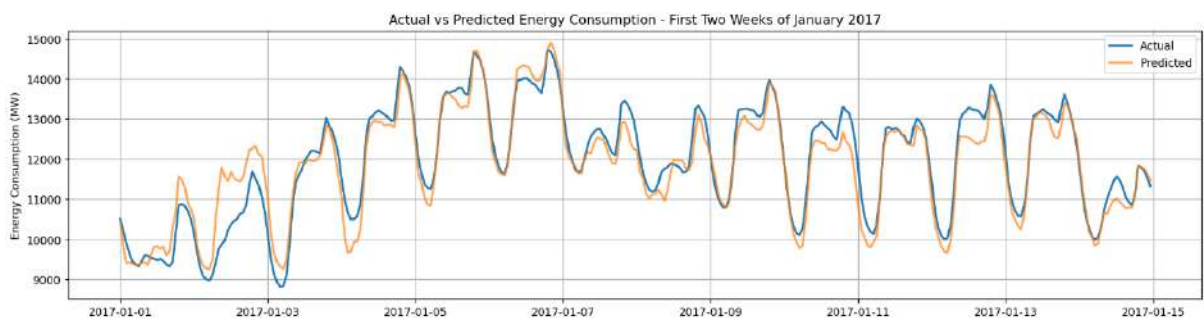


Рис. 12. Результати роботи LightGBM.

2.2.4. Прогнозування енергоспоживання з використанням нейронних мереж.

Нейронні мережі являють собою потужний інструмент машинного навчання, здатний моделювати складні нелінійні залежності в даних енергоспоживання будівель. На відміну від класичних статистичних методів, нейронні мережі не потребують явного визначення математичної моделі взаємозв'язків між змінними, що робить їх особливо цінними при аналізі енергетичних часових рядів зі складною структурою та прихованими закономірностями.

Архітектури рекурентних нейронних мереж (RNN) спеціально розроблені для моделювання послідовних даних, де поточний стан залежить від попередніх значень, що робить їх ідеальними для аналізу часових рядів енергоспоживання.

Математично функціонування базової RNN можна описати рівняннями:

$$h_t = \sigma(W_h \cdot h_{t-1} + W_x \cdot x_t + b_h) \quad y_t = W_y \cdot h_t + b_y, \text{ де:}$$

h_t — прихований стан мережі в момент часу t

x_t — вхідний вектор у момент часу t

y_t — вихідний вектор у момент часу t

W_h, W_x, W_y — матриці вагових коефіцієнтів

b_h, b_y — вектори зміщення

σ — функція активації

Однак класичні RNN стикаються з проблемами зникаючого або вибухового градієнта при навчанні на довгих послідовностях, що обмежує їх здатність виявляти довгострокові залежності, характерні для енергоспоживання будівель.

LSTM мережі для моделювання довгострокових залежностей

Long Short-Term Memory (LSTM) мережі були розроблені як удосконалення класичних RNN, спрямоване на вирішення проблеми зникаючого градієнта та покращення здатності моделювати довгострокові залежності в даних.

Архітектура LSTM включає спеціальні комірки пам'яті з механізмами воріт (gates), які контролюють потік інформації:

- Вхідні ворота (input gate) — вирішують, яка нова інформація буде збережена в стані комірки: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$ $\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$
- Ворота забування (forget gate) — визначають, яка інформація залишиться в стані комірки: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
- Оновлення стану комірки: $C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$
- Вихідні ворота (output gate) — контролюють, яка інформація буде передана на вихід: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$ $h_t = o_t \cdot \tanh(C_t)$

Ця архітектура дозволяє LSTM мережам ефективно моделювати як короткострокові, так і довгострокові залежності в енергоспоживанні будівель, зокрема добові цикли активності, тижневі патерни, сезонні коливання, пов'язані з погодними умовами та річні тренди.

Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) мережі представляють собою спрощену версію LSTM, що зберігає більшість переваг при зменшенні обчислювальної складності. GRU об'єднує ворота забування та вхідні ворота в єдині "ворота оновлення" (update gate), а також має "ворота скидання" (reset gate).

Переваги GRU для прогнозування енергоспоживання: менша кількість параметрів порівняно з LSTM, що знижує ризик перенавчання; швидше навчання та виведення прогнозів; ефективність при обмежених обсягах навчальних даних

Обидві архітектури мають свої переваги в контексті моделювання енергоспоживання будівель:

Характеристика	GRU	LSTM
Складність моделі	Нижча	Вища
Обчислювальна ефективність	Вища	Нижча

Моделювання дуже довгих залежностей	Достатньо ефективно для більшості завдань	Дещо краще завдяки явній комірці пам'яті
Стійкість до перенавчання	Природно більш стійка через меншу кількість параметрів	Потребує більшої регуляризації
Обсяг необхідних навчальних даних	Менший	Більший

Табл. 1. Порівняльний аналіз GRU та LSTM моделей.

У практичних задачах прогнозування енергоспоживання вибір між LSTM та GRU часто визначається експериментальним шляхом на основі крос-валідації.

2.3 Проблематика статичних моделей прогнозування енергоспоживання будівель.

У даному дослідженні було проведено аналіз великого масиву даних енергоспоживання будівель з метою виявлення можливості прогнозування показників споживання енергії, спираючись винятково на набір вхідних параметрів без урахування історичних даних часового ряду.

Використаний датасет містив 38 параметрів, включаючи тип будівлі, температуру, вологість, заповнюваність, споживання енергії різними системами, ціну енергії, віковий стан будівлі та обладнання, енергоефективність, розмір будівлі та інші показники. Попри значну кількість факторів, спроби побудови регресійних моделей для прогнозування загального енергоспоживання виявились неефективними.

Результати аналізу кореляційних зв'язків

Проведений аналіз кореляційної матриці виявив відсутність значущих лінійних взаємозв'язків між енергоспоживанням будівель та вхідними параметрами. Як видно з наведеної кореляційної матриці та коефіцієнтів кореляції, найвищі значення кореляції з цільовою змінною "Energy Consumption (kWh)" є надзвичайно низькими:

1. Voltage Levels (V): 0.008765
2. Water Usage (liters): 0.008484
3. Energy Savings Potential (%): 0.008018
4. Grid Stability Score: 0.005343
5. IoT Sensor Count: 0.004746

Ці показники знаходяться близько до нуля, що свідчить про практично відсутній лінійний зв'язок між енергоспоживанням та даними параметрами. Особливо примітно, що навіть показники HVAC Consumption (kWh), які логічно мали б демонструвати сильну кореляцію з загальним енергоспоживанням як його компонент, показують лише мінімальний кореляційний зв'язок (0.002832).

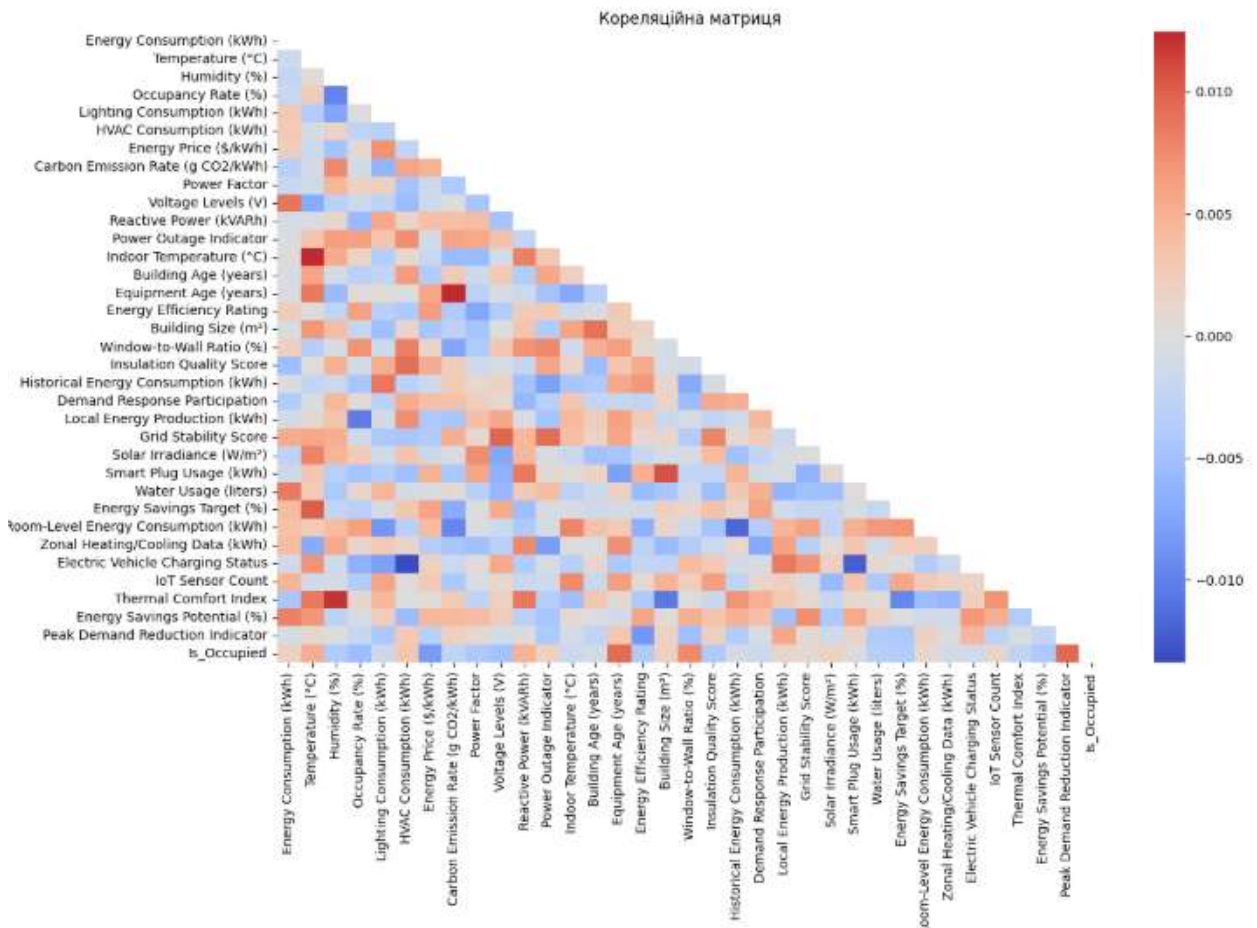


Рис. 13. кореляційна матриця вхідних параметрів.

Аналіз важливості ознак для моделей "Випадковий ліс" та "Градiєнтний бустинг" показав, що найбільш впливові параметри включають:

1. Insulation Quality Score

2. Equipment Age (years)
3. Temperature (°C)
4. Building Size (m²)
5. Occupancy Rate (%)

Проте ці параметри виявляють доволі обмежену ефективність щодо точності передбачень, що підтверджується недостатніми загальними показниками ефективності розроблених моделей.

Проведений аналіз яскраво ілюструє принципову недостатність методології прогнозування енерговитрат споруд, що спирається винятково на незмінні характеристики, ігноруючи темпоральні послідовності даних. Ключові узагальнення дослідження можна окреслити так:

Дослідження показало критично слабкі кореляції між енергоспоживанням та всіма досліджуваними факторами, що робить неможливим створення ефективної лінійної моделі прогнозування. Фактичні дані переконливо свідчать, що енергоспоживання будівель є комплексним явищем з нелінійними залежностями, часовими трендами та циклічністю, які вимагають застосування більш складного аналізу часових рядів замість спрощених лінійних підходів. Ці результати практично продемонстрували ті особливості в порівнянні дискретних моделей та моделей, побудованих на часових рядах, про які розповідалося у дослідженні [13], де серед переваг першої було названо її простоту і можливість уникнення стадії навчання, але так само вказано проблематику низької точності та важкості моделювання реальних сценаріїв.

Саме тому прогностичні моделі застосовують часові зсуви (time lags). Часові зсуви представляють значення показників не в теперішній момент, а в попередніх періодах. У сфері прогнозування енергоспоживання будівель, впровадження часових зсувів передбачає інтеграцію архівних даних щодо енерговикористання за попередні години, доби, тижні чи місяці як предикативних чинників для формування прогнозів майбутнього споживання.

2.4. Вхідні дані та аналіз отриманих результатів.

2.4.1. Обрана вибірка даних.

Для проведення експерименту було обрано вибірку даних з м. Чикаго. Датасет представляє собою комплексну часову серію для аналізу та прогнозування енергоспоживання в місті Чикаго. Вихідна вибірка містить 66489 погодинних записів, починаючи з 31 грудня 2011 року, і була значно розширена додатковими факторами для підвищення аналітичної цінності. Окрім основної метрики споживання електроенергії в мегаватах (COMED_MW), датасет збагачено кліматичними показниками з аеропорту Чикаго: температурою (Chicago_temp, у Кельвінах), вологістю (Chicago_humidity, у відсотках), атмосферним тиском (Chicago_pressure), напрямком (Chicago_wind_dir) та швидкістю вітру (Chicago_wind_speed). Для розширення можливостей прогнозування моделями машинного навчання також було додано часові характеристики: година доби (hour), день тижня (weekday) та індикатор святкового дня (is_holiday).

2.4.2. Результати статистичних моделей.

При порівнянні результатів, варто зазначити, що кожен наступний метод досягає певного покращення у прогнозуванні, проте для більш змістовної оцінки варто звернутися до метрик. MAE (Mean Absolute Error) — це середнє абсолютне значення помилок між прогнозованими та фактичними значеннями. При використанні мін-макс функції для нормалізації даних, фактично, MAE перетворюється у відсоткове значення відхилення. MAPE (Mean Absolute Percentage Error) — це середній відсоток помилки між прогнозом і реальністю. Він показує, на скільки відсотків у середньому модель помиляється.

Представлені дані показують чітку еволюцію точності прогнозування від простіших до більш складних часових моделей. На основі показника MAPE можна побачити послідовне підвищення якості прогнозів: ARIMA демонструє найвищу похибку в 14.88%, SARIMA покращує результат до 12.56%, SARIMAX далі знижує похибку до 11.8019%, а найкращий результат дає SARIMAX з кліматичними параметрами - 11.2309%. Це свідчить про значне

покращення прогнозів при врахуванні сезонності та особливо при включенні екзогенних змінних, зокрема кліматичних параметрів.

Варто відзначити зв'язок між кількістю параметрів моделей та їхньою ефективністю. ARIMA і SARIMA обидві використовують лише 2 параметри, але врахування сезонності в SARIMA дає відчутне покращення. SARIMAX з 5 параметрами демонструє значно кращі результати в усіх сезонах, а найскладніша модель SARIMAX з кліматичними параметрами використовує 16 параметрів і досягає найвищої точності. Це свідчить, що збільшення складності моделі через введення релевантних параметрів суттєво покращує прогнозування.

	ARIMA	SARIMA	SARIMAX	SARIMAX with climate parameteres
Кількість параметрів	2	2	5	16
MAPE	14.88%	12.56%	11.8019%	11.2309%
Summer RMSE	2992.25 (14.69%)	2760.42 (13.55%)	2576.9607 (12.65%)	2354.1294 (11.56%)
Spring RMSE	1850.54 (9.08%)	1450.03 (7.12%)	1509.4182 (7.41%)	1109.4182 (5.45%)
Winter RMSE	1391.59 (6.83%)	1180.93 (5.80%)	1085.2161 (5.33%)	1058.1728 (5.19%)

Табл. 2. Порівняльний аналіз різних статистичних моделей.

2.4.3. Результати моделей з використанням машинного навчання.

Аналіз представлених результатів дослідження різних моделей прогнозування енергоспоживання виявляє цікаві закономірності та демонструє чіткі відмінності в ефективності алгоритмів. Загальна точність прогнозування найкраще відображається через коефіцієнт детермінації R^2 , де XGBoost

лідирує з показником 0.8846, що свідчить про його здатність пояснювати близько 88.5% варіації в даних енергоспоживання. LightGBM та Random Forest також демонструють високу точність з показниками 0.8705 та 0.8611 відповідно. Нейромережові підходи представлені LSTM з R^2 на рівні 0.8293 та GRU з 0.7935, що хоч і нижче за лідерів, але все ж значно краще за класичну лінійну регресію з її 0.7012.

Особливо примітна сезонна варіативність у прогнозуванні. Літній період становить найбільший виклик для всіх моделей, де навіть лідер XGBoost має RMSE 621.46 (4.93%), тоді як лінійна регресія показує значно гіршу точність з RMSE 1422.34 (11.29%). Весняний період прогнозується найкраще більшістю моделей, причому XGBoost знову лідирує з RMSE 506.96 (4.89%), а LightGBM посідає друге місце з показником 542.75 (5.24%). Зимові прогнози демонструють середню складність для всіх алгоритмів, і навіть тут градієнтний бустинг показує перевагу.

Попри однакову кількість параметрів у всіх моделях (31), що забезпечує коректність порівняння, спостерігається виразна перевага ансамблевих методів машинного навчання над рекурентними нейронними мережами та класичними статистичними підходами. XGBoost не лише демонструє найвище значення R^2 , але й найменшу похибку RMSE в усіх сезонах, що вказує на його стабільність і універсальність для задачі прогнозування енергоспоживання. Тоді як RF та LightGBM за певних умов наближаються до лідера, рекурентні нейронні мережі (GRU та LSTM) показують кращі результати порівняно з лінійною регресією, але поступаються ансамблевим методам як у загальній точності, так і в сезонній стабільності. Цей результат є відмінним від попередніх досліджень, так, наприклад у роботі [6] було досягнуто кращих показників прогнозування саме в ШНМ. Цікаво, що хоча LSTM загалом перевершує GRU за коефіцієнтом R^2 , вона демонструє гіршу точність улітку, що може свідчити про її чутливість до специфічних літніх патернів енергоспоживання.

	Linear Regression	RF	XGBoost	LightGBM	GRU	LSTM
R ²	0.7012	0.8611	0.8846	0.8705	0.7935	0.8293
Summer RMSE	1422.34 (11.29%)	983.60 (7.81%)	621.46 (4.93%)	949.38 (7.54%)	1357.23 (10.78%)	1214.23 (9.64%)
Spring RMSE	897.15 (8.66%)	628.11 (6.06%)	506.96 (4.89%)	542.75 (5.24%)	661.64 (6.39%)	612.41 (5.91%)
Winter RMSE	1011.26 (8.66%)	753.17 (6.45%)	739.56 (6.33%)	795.20 (6.81%)	815.10 (6.98%)	762.27 (6.53%)

Табл. 3. Порівняльний аналіз моделей машинного навчання.

Розділ 3. Експериментальна частина.

У рамках експериментальної частини дослідження було розроблено та випробувано гібридну модель для прогнозування енергоспоживання, що поєднує переваги двох потужних методів машинного навчання: XGBoost та LSTM. Перша модель змогла продемонструвати найкращі результати прогнозування, у той час як друга - здатна покращити результати зарахунок здатності обробляти часові послідовності та виявляти довгострокові залежності в даних.

3.1 Будова створеної гібридної моделі.

У процесі розробки запропонованої гібридної моделі прогнозування було поєднано переваги нейронних мереж та градієнтного бустингу. Двонаправлений LSTM (Long Short-Term Memory) забезпечує здатність моделі враховувати історичні дані як з попередніх, так і з наступних часових кроків, що дає змогу моделі формувати більш повне розуміння динаміки часового ряду. На відміну від класичних однонаправлених LSTM, які опрацьовують дані лише в прямому хронологічному порядку, двонаправлена архітектура (рис. 14) забезпечує одночасний аналіз часової послідовності в обох напрямках, що істотно збагачує контекстуальне представлення кожного часового кроку. [14]

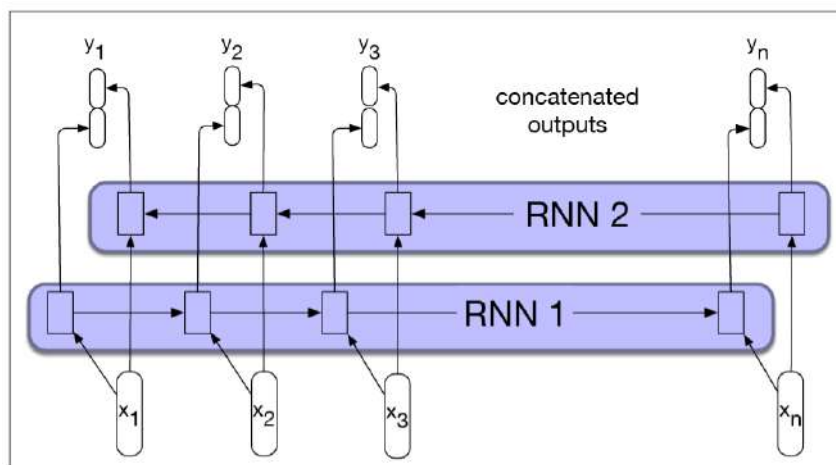


Рис. 14. Внутрішня будова двонаправлених рекурентних нейронних моделей. [14]

Серед інших удосконалень варто зазначити механізм уваги [15], який надає LSTM моделі можливість динамічно розподіляти ваги важливості між різними часовими кроками залежно від їхньої релевантності для конкретного прогнозу. Механізм уваги обчислює вагові коефіцієнти для кожного прихованого стану LSTM, визначаючи, на які часові періоди модель повинна звернути більшу увагу при формуванні прогнозу.

Для запобігання перенавчанню, що є типовою проблемою для складних нейромережових архітектур, особливо при обмеженому обсязі навчальних даних, було впроваджено регуляризацію L2-рівня. Цей метод регуляризації додає до функції втрат додатковий штрафний член, пропорційний квадрату норми векторів ваг, що спонукає модель віддавати перевагу меншим за абсолютною величиною вагам і тим самим зменшує складність моделі. [14]

Особливо корисним була реалізація механізму ранньої зупинки, який моніторить продуктивність моделі на валідаційному наборі даних і припиняє навчання, коли показники якості на цьому наборі перестають покращуватися, запобігаючи тим самим надмірному підлаштуванню моделі під особливості навчальних даних. Окрім цього, це дало змогу швидше тестувати зміни у кодї, адже моделі відпрацьовували швидше через коротший цикл роботи.

Паралельно з покращеннями LSTM моделі було реалізовано оптимізаційний алгоритм рою частинок (Particle Swarm Optimization, PSO).

PSO представляє собою метаевристичний алгоритм, який у контексті налаштування XGBoost, управляв популяцією потенційних рішень (наборів гіперпараметрів), де кожна "частинка" представляла окрему конфігурацію моделі. Частинки переміщувалися в просторі пошуку, керуючись як власним досвідом, так і колективним знанням всієї популяції, що дозволяло ефективно досліджувати різноманітні комбінації параметрів та поступово збігатися до оптимальних рішень. [16]

Для оцінки якості кожної потенційної конфігурації XGBoost використовувалася цільова функція на основі середньоквадратичної помилки (Mean Squared Error, MSE), що обчислює середній квадрат різниці між прогнозованими та фактичними значеннями. Процес оптимізації включав автоматичний пошук найкращої комбінації таких гіперпараметрів як глибина дерев, швидкість навчання, коефіцієнт регуляризації, мінімальна вага спостережень у листовому вузлі та інших, що визначають структуру та поведінку алгоритму градієнтного бустингу. Фінальне навчання моделі XGBoost здійснювалося з використанням знайдених оптимальних параметрів.

Для ефективного об'єднання прогнозів окремих моделей було реалізовано мета-рівень на основі Ridge регресії [14], яка належить до класу лінійних регресійних моделей з L2-регуляризацією. У цьому підході прогнози, генеровані LSTM та XGBoost моделями, слугували вхідними даними для мета-моделі, яка навчалася визначати оптимальні вагові коефіцієнти для кожного з базових прогнозів. Ridge регресія визначає ці ваги, мінімізуючи суму квадратів похибок прогнозу з додатковим регуляризаційним членом [17]. Цей підхід динамічно адаптує відносний внесок кожної базової моделі, що забезпечує здатність гібридної моделі виявляти та використовувати взаємодоповнюючі сильні сторони різних прогностичних підходів, а отже підвищувати загальну продуктивність системи порівняно з окремими її компонентами.

3.2 Отримані результати.

Результати експериментів підтвердили значну перевагу гібридного підходу порівняно з окремими моделями. Після всіх покращень LSTM модель показала такі результати: $RMSE = 702.81 \text{ MW}$ (6.12%), $MAE = 482.43 \text{ MW}$ (4.20%), $R^2 = 0.8768$. XGBoost модель продемонструвала ще кращі результати: $RMSE = 647.28 \text{ MW}$ (5.63%), $MAE = 443.16 \text{ MW}$ (3.86%), $R^2 = 0.8955$. Бачимо, що введені покращення підвищили загальну ефективність кожної моделі окремо. Побудована ж на основі моделей ridge-регресія продемонструвала найкращі фінальні показники: $RMSE = 615.36 \text{ MW}$ (5.35%), $MAE = 424.82 \text{ MW}$ (3.70%), $R^2 = 0.9055$. Гібридна модель забезпечила покращення всіх метрик, зокрема зменшення загальної $RMSE$ на 5% порівняно з XGBoost та на 12.4% порівняно з LSTM. Також покращилися сезонні показники: літня $RMSE$ знизилась на 5% (з 624.32 MW (5.43%) до 593.10 MW (5.16%)) порівняно з XGBoost, весняна $RMSE$ – на 5% (з 658.47 MW (5.73%) до 625.55 MW (5.44%)) порівняно з XGBoost та на 11.3% порівняно з LightGBM, зимова $RMSE$ – на 6.7% (з 712.41 MW (6.20%) до 664.68 MW (5.78%)) порівняно з Random Forest моделлю.

Цікавим результатом дослідження, є аналіз сезонної ефективності, що показав, що різні алгоритми працюють краще в різні пори року. Саме тому комбінований підхід дає змогу компенсувати слабкі сторони окремих моделей.

Для подальшого дослідження варто розглянути можливість розширення вхідних даних та їхнього спектру та поліпшення методології дослідження. У питанні розширення вхідних даних, було б корисно мати таку інформацію:

1. Додаткові метеорологічні параметри, такі як опади, хмарність, сонячна радіація, точка роси, видимість.
2. Дані з кількох метеостанцій регіону, а не лише з Чикаго, що дасть змогу врахувати географію погодних змін та, як наслідок, передбачити можливі погодні зміни вже для нашого міста.
3. Економічні показники, такі як індекси промислової активності, дані про зайнятість, ціни на електроенергію, що можуть впливати на споживання.

4. Історичні дані про відключення та технічні проблеми в мережі, що можуть пояснити аномальні значення у ретроспективних даних, а отже зробити вхідний масив даних більш очищеним та, як наслідок, зменшити вплив викидів.
5. Часові ряди споживання електроенергії за секторами (промисловість, комерційний, житловий), що дозволить точніше моделювати різні компоненти навантаження, адже як було зазначено в другому розділі, кожний сектор має власний енергетичний профіль споживання.

Методологічні вдосконалення можуть включати розробку окремих моделей для різних типів днів та сезонів, можливо навіть використання архітектури трансформерів, які добре працюють з довгими послідовностями та можуть виявляти складні залежності в даних.

3.3 Виклики прогнозування енергоспоживання будівель в контексті України.

Критичним фактором, що суттєво знижує ефективність прогнозування, є підвищена волатильність енергосистеми, спричинена постійними обстрілами стратегічних об'єктів енергетичної інфраструктури. Цілеспрямовані атаки на генеруючі потужності, трансформаторні підстанції, об'єкти генерації та розподілу електроенергії призводять до суттєвих обмежень доступної потужності, що зумовлює необхідність оперативного коригування параметрів генерації та споживання. В таких умовах класичні моделі прогнозування, що базуються на історичних даних, втрачають свою актуальність або демонструють низьку точність.

На особливу методологічну увагу заслуговує аналіз графіків аварійних або стабілізаційних відключень, які впроваджуються при виникненні значного дефіциту потужності. Їх реалізація, хоча і є виправданою з точки зору підтримання стабільності енергосистеми, генерує високу волатильність системи та низьку передбачуваність моделей поведінки споживачів, особливо в житловому та комерційному секторах.

Крім того, у воєнний час значно знижується достовірність вхідних даних. Багато лічильників або не передають інформацію вчасно, або виходять з ладу. У деяких регіонах втрачені цілі вузли обліку, особливо в районах активних бойових дій або на окупованих територіях.

Однак, незважаючи на всі труднощі, важливість точного прогнозування, безумовно, зростає. Адже своєчасне та обґрунтоване прогнозування рівнів споживання дозволяє розробляти більш чіткі графіки відключень, які є більш справедливими та менш стресовими для споживачів; підвищувати ефективність реагування на втрати потужності, запобігаючи розбалансуванню системи; оптимізувати розподіл резервів, генеруючих та диспетчерських ресурсів; впроваджувати алгоритми адаптивного управління навантаженням, зокрема для об'єктів критичної інфраструктури.

На жаль, проведення повноцінних наукових досліджень та побудова адекватних моделей прогнозування в нинішніх умовах неможливі, оскільки відсутній відкритий доступ до актуальних даних про споживання, відключення, пошкодження об'єктів енергосистеми та пов'язані з ними оперативні рішення. Більшість відповідної інформації є засекреченою з міркувань безпеки, що є виправданим в умовах війни, але суттєво обмежує аналітичні та прогнозні можливості.

Таким чином, розробка ефективних інструментів прогнозування енергоспоживання у воєнний час залишається стратегічним завданням, але його реалізацію варто розглядати після закінчення бойових дій або за сприяння держави та розробки відповідного програмного забезпечення в закритих умовах.

Висновки

У ході виконання курсової роботи було проведено всебічне дослідження методів прогнозування енергоспоживання будівель, що базуються на аналізі часових рядів різної періодичності. Проаналізовано особливості денного, тижневого та річного часових рядів енергоспоживання; розглянуто та реалізовано різні моделі прогнозування - від класичних статистичних методів (ARIMA, експоненційне згладжування) до сучасних алгоритмів машинного навчання (лінійна регресія, Random Forest, XGBoost) та нейронних мереж.

Порівняння цих методів дало змогу виявити їхні переваги та недоліки при застосуванні до даних енергоспоживання будівель. Встановлено, що статистичні методи добре працюють на стабільних часових рядах з чітко вираженою сезонністю, проте мають обмеження при аналізі складних нелінійних залежностей. Алгоритми машинного навчання показали кращі результати при роботі з багатовимірними даними, де важливо враховувати додаткові фактори впливу, такі як погодні умови, день тижня чи особливості експлуатації будівлі.

Найбільш точні результати прогнозування було отримано за допомогою ансамблевих методів на основі дерев рішень (XGBoost), які забезпечили найменшу похибку в порівнянні з іншими розглянутими підходами. Нейронні мережі показали високу адаптивність, проте вимагали значно більше обчислювальних ресурсів та часу на навчання.

У рамках експериментальної частини дослідження було розроблено та випробувано гібридну модель для прогнозування енергоспоживання. Завдяки оптимізаційному алгоритму PSO та об'єднанню результатів окремих моделей, було досягнуто середньої точності прогнозування в понад 90% (R^2).

Посилання та використана література

1. Економічна правда. Відповідальне енергоспоживання в умовах війни: як вирішують проблему лічильники Sempal. *Економічна правда*. URL: <https://epravda.com.ua/publications/2024/09/25/719689/>
2. Renewable Energy Market Size to Reach USD 7.28 Trillion by 2034. *Precedence Research - Market Research Reports & Strategic Consulting*. URL: <https://www.precedenceresearch.com/renewable-energy-market>
3. Machine learning for energy performance prediction at the design stage of buildings / R. Olu-Ajayi et al. *Energy for Sustainable Development*. 2022. Vol. 66. P. 12–25. URL: <https://doi.org/10.1016/j.esd.2021.11.002>
4. Advancing building energy efficiency: a deep learning approach to early-stage prediction of residential electric consumption / K. Sundaram et al. *Energy reports*. 2024. Vol. 12. P. 1281–1292. URL: <https://doi.org/10.1016/j.egy.2024.07.034>
5. Predicting energy consumption of residential buildings using metaheuristic-optimized artificial neural network technique in early design stage / M. R. Kaloop et al. *Building and environment*. 2025. P. 112749. URL: <https://doi.org/10.1016/j.buildenv.2025.112749>
6. Mohan R., Pachauri N. An ensemble model for the energy consumption prediction of residential buildings. *Energy*. 2025. Vol. 314. P. 134255. URL: <https://doi.org/10.1016/j.energy.2024.134255>
7. A quantitative comparison of different methods for air conditioning energy consumption prediction in residential buildings / C. Du et al. *Energy and built environment*. 2024. URL: <https://doi.org/10.1016/j.enbenv.2024.11.007>
8. Borodinecs A., Lebedeva K., Odineca T. Evaluation of household electricity consumption in multi-apartment buildings for optimization of rooftop PV systems. *Energy and buildings*. 2024. P. 114971. URL: <https://doi.org/10.1016/j.enbuild.2024.114971>
9. Xiao S. A novel improved model for green building energy consumption prediction based on time-series analysis. *Proceedings of the institution of civil*

engineers - smart infrastructure and construct
<https://us05web.zoom.us/j/82543431282?pwd=r72CHSxUGOuYtQWbrcQwTUGxpDt3bN.Ion>. 2023. P. 1–10. URL: <https://doi.org/10.1680/jsmic.22.00028>

10. А. Т. Яровий, Є. М. Страхов. Аналіз часових рядів. — Одеса : Освіта України, 2019. — 109 с.

11. Ahmad M. W., Mourshed M., Rezgui Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*. 2017. Vol. 147. P. 77–89. URL: <https://doi.org/10.1016/j.enbuild.2017.04.038>

12. Berardi U. A cross-country comparison of the building energy consumptions and their trends. *Resources, Conservation and Recycling*. 2017. Vol. 123. P. 230–241. URL: <https://doi.org/10.1016/j.resconrec.2016.03.014>

13. A review on time series forecasting techniques for building energy consumption / C. Deb et al. *Renewable and Sustainable Energy Reviews*. 2017. Vol. 74. P. 902–924. URL: <https://doi.org/10.1016/j.rser.2017.02.085>

14. Speech and Language Processing. *Stanford University*. URL: <https://web.stanford.edu/~jurafsky/slp3/>

15. 1. Smith A. W., Zipser D. LEARNING SEQUENTIAL STRUCTURE WITH THE REAL-TIME RECURRENT LEARNING ALGORITHM. *International Journal of Neural Systems*. 1989. Vol. 01, no. 02. P. 125–131. URL: <https://doi.org/10.1142/s0129065789000037>

16. Gad A. G. Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review. *Archives of Computational Methods in Engineering*. 2022. URL: <https://doi.org/10.1007/s11831-021-09694-4>

17. Ridge. *scikit-learn*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html