

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ПРИРОДОКОРИСТУВАННЯ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ЗАОЧНОЇ ТА
ПІСЛЯДИПЛОМНОЇ ОСВІТИ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

КВАЛІФІКАЦІЙНА РОБОТА
другого (магістерського) рівня вищої освіти

на тему: «Дослідження моделей машинного навчання для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками»

Виконав: студент групи Іт-61

Спеціальності 126 «Інформаційні системи та технології»

(шифр і назва)

Білецький Роман Романович

(Прізвище та ініціали)

Керівник: к.т.н., в.о. доцента Татомир А.В.

(Прізвище та ініціали)

Рецензент: к.т.н., доцент Сиротюк С.В.

(Прізвище та ініціали)

ДУБЛЯНИ-2022

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ПРИРОДОКОРИСТУВАННЯ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ЗАОЧНОЇ ТА
ПІСЛЯДИПЛОМНОЇ ОСВІТИ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Другий (магістерський) рівень вищої освіти
Спеціальність 126 «Інформаційні системи та технології»

«ЗАТВЕРДЖУЮ»

Завідувач кафедри _____

д.т.н., проф. А.М. Тригуба

« ____ » _____ 2022 р.

ЗАВДАННЯ

на кваліфікаційну роботу студенту

Білецькому Роману Романовичу

1. Тема роботи: «Дослідження моделей машинного навчання для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками»

Керівник роботи Татомир Андрій Володимирович, доцент
затверджені наказом по університету від 30.06.2022 року №137/к-с.

2. Строк подання студентом роботи 10.12.2022 р.

3. Вихідні дані до роботи: алгоритми машинного навчання; методика дослідження моделей машинного навчання; методика прогнозування алгоритмами машинного навчання із використанням інтелектуального аналізу даних.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які необхідно розробити) _____

Вступ.

1. Аналіз стану предметної галузі та завдання кваліфікаційної роботи.

2. Моделі прогнозування із використанням методів машинного навчання.

3. Результати дослідження моделей машинного навчання для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками;

4. Охорона праці та безпека у надзвичайних ситуаціях.

5. Оцінення ефективності запропонованої моделі.

Висновки та пропозиції.

Список використаної літератури.

5. Перелік ілюстраційного матеріалу (з точним зазначенням обов'язкових слайдів): аналіз стану споживання електроенергії будівлями; аналіз існуючих методів машинного навчання; вибір методів машинного навчання для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками; результати наближеного оцінення моделей прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками; результати підготовки та аналіз даних для прогнозування; результати обґрунтування раціональної моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками; економічна ефективність.

6. Консультанти з розділів:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1, 2, 3, 5	<i>Татомир А.В., в.о. доцента кафедри інформаційних технологій</i>		
4	<i>Городецький І.М., доцент кафедри управління проектами та безпеки виробництва</i>		

7. Дата видачі завдання

30 червня 2022 р.

Календарний план

№ з/п	Назва етапів кваліфікаційної роботи	Терміни виконання етапів роботи	Примітка
1	<i>Написання першого розділу</i>	<i>30.06-04.07.22</i>	
2	<i>Виконання другого розділу та аркушів ілюстраційного матеріалу до нього</i>	<i>05.07-14.08.22</i>	
3.	<i>Виконання третього розділу та аркушів ілюстраційного матеріалу до нього</i>	<i>15.08-24.09.22</i>	
4.	<i>Написання розділу «Охорона праці та безпека у надзвичайних ситуаціях»</i>	<i>25.09-10.10.22</i>	
5.	<i>Оцінення ефективності запропонованої системи</i>	<i>20.10-31.10.22</i>	
6.	<i>Завершення оформлення розрахунково-пояснювальної записки та аркушів ілюстраційного матеріалу</i>	<i>01-30.11.22</i>	
7.	<i>Завершення роботи в цілому</i>	<i>01-10.12.22</i>	

Студент _____ Білецький Р.Р.
(підпис)

Керівник роботи _____ Татомир А.В.
(підпис)

УДК 621.311.1

Дослідження моделей машинного навчання для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.

Білецький Р.Р. Кафедра інформаційних технологій – Дубляни, Львівський НУП, 2022.

Кваліфікаційна робота: 81 с. текст. част., 18 рис., 2 табл., 10 арк. ілюстраційного матеріалу, 44 джерела.

Виконано аналіз стану споживання електроенергії будівлями та обґрунтування доцільності її прогнозування. Подано методи прогнозування обсягів генерування та споживання електроенергії. Наведена класифікація методів машинного навчання. Проаналізовано особливості оцінки енергетичної ефективності будівель за допомогою засобів машинного навчання. Сформульовано завдання кваліфікаційної роботи.

Здійнено вибір методів машинного навчання для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками. Проаналізовано метод дерев рішень (Decision Tree Regressor), регресію випадкового лісу, багатосаровий перцептрон (MLP Regressor), підвищення градієнта (Gradient Boosting Regressor).

Здійснено підготовку та аналіз даних для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками. Проведено оцінення взаємозв'язків між даними для прогнозування. Здійснено вибір інструментарію та підготовка даних до машинного навчання. Проведено наближене оцінення моделей прогнозування. Подано результати обґрунтування раціональної моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.

Розроблено заходи із охорони праці та безпека у надзвичайних ситуаціях. Визначено ефективність від використання моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.

ЗМІСТ

ВСТУП.....	7	
АНАЛІЗ СТАНУ ПРЕДМЕТНОЇ ГАЛУЗІ ТА ЗАВДАННЯ		
КВАЛІФІКАЦІЙНОЇ РОБОТИ.....	8	
1.1. Аналіз стану споживання електроенергії будівлями та обґрунтування доцільності її прогнозування	9	
1.2. Методи прогнозування обсягів генерування та споживання електроенергії.....	12	
1.3. Класифікація методів машинного навчання.....	14	
1.4. Оцінка енергетичної ефективності будівель за допомогою засобів машинного навчання.....	17	
1.5. Завдання кваліфікаційної роботи.....	19	
2. ВИБІР МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ		
ПРОГНОЗУВАННЯ ОБСЯГІВ ГЕНЕРУВАННЯ ТА СПОЖИВАННЯ		
ЕЛЕКТРОЕНЕРГІЇ ІНДИВІДУАЛЬНИМИ БУДИНКАМИ.....		20
2.1. Метод дерев рішень (Decision Tree Regressor)	20	
2.2. Регресія випадкового лісу (Random Forest Regressor).....	25	
2.3. Багатошаровий перцептрон (MLP Regressor).....	27	
2.4. Підвищення градієнта (Gradient Boosting Regressor).....	29	
3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ МОДЕЛЕЙ МАШИННОГО		
НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ОБСЯГІВ ГЕНЕРУВАННЯ ТА		
СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ ІНДИВІДУАЛЬНИМИ БУДИНКАМИ...37		
3.1. Підготовка та аналіз даних для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками	37	
3.2. Оцінення взаємозв'язків між даними для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.....	49	
3.3. Вибір інструментарію та підготовка даних до машинного навчання.....	53	
3.4. Наближене оцінення моделей прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.....	55	

3.5. Результати обґрунтування раціональної моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.....	57
4. ОХОРОНА ПРАЦІ ТА БЕЗПЕКА У НАДЗВИЧАЙНИХ СИТУАЦІЯХ	63
4.1. Аналіз небезпечних та шкідливих виробничих чинників під час роботи з комп'ютерною технікою	63
4.2. Моделювання процесу виникнення травм та аварій	65
4.3. Розробка заходів щодо безпеки у надзвичайних ситуаціях	67
6. ВИЗНАЧЕННЯ ЕФЕКТИВНОСТІ ВІД ВИКОРИСТАННЯ МОДЕЛІ ПРОГНОЗУВАННЯ ОБСЯГІВ ГЕНЕРУВАННЯ ТА СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ ІНДИВІДУАЛЬНИМИ БУДИНКАМИ	69
ВИСНОВКИ І ПРОПОЗИЦІЇ.....	72
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	77

ВСТУП

У період війни на території України, де частково пошкоджено енергетичні системи, досить актуальними є задачі планування енергоспоживання як окремими підприємствами, так і індивідуальними будинками. Під час виконання планування виникають задачі щодо точного прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками. Кількісні значення обсягів генерування та споживання електроенергії індивідуальними будинками значною мірою залежать від множини чинників, які мають системні взаємозв'язки. Ці взаємозв'язки між зазначеними чинниками у більшості випадків є нелінійними і їх досить тяжко описати певними алгоритмами.

На даний час у літературі є описано велику кількість підходів, методів та інструментарію як для планування енергоспоживання, так і для прогнозування обсягів генерування [5; 12]. Вони базуються, як на алгоритмічних підходах, так і передбачають використання сучасних технологій – машинного навчання.

Розв'язання науково-прикладної задачі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками потребує проведення досліджень щодо доцільності використання методів машинного навчання. Для цього існує потреба виконання аналізу відомих алгоритмів та результатів проведених досліджень у галузі машинного навчання, які стосуються коротко строкового прогнозування окремих подій, які змінюються у часі.

Під час виконання досліджень із використанням моделей машинного навчання існує потреба у використанні історичних даних щодо обсягів споживання електроенергії та природно-кліматичних умов на території розташування індивідуальних будинків. Для цього можна скористатися даними, які наявні на сервісах у відкритому доступі. Зокрема, це стосується даних щодо тенденцій зміни температури навколишнього середовища та вологості повітря,

а також зібраних статистичних даних щодо обсягів споживання електроенергії індивідуальними будинками.

На підставі вище сказаного можна вважати, що тема кваліфікаційної роботи «Дослідження моделей машинного навчання для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками» є досить актуальною та своєчасною, а також має практичну цінність.

Об'єктом дослідження є методи машинного навчання, які забезпечують прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками на підставі історичних даних.

Предмет дослідження є взаємозв'язки між обсягами генерування та споживання електроенергії індивідуальними будинками та природно-кліматичними умовами їх розташування.

Метою роботи є підвищення ефективності прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками завдяки обґрунтуванню раціональної моделі, що базується на алгоритмі машинного навчання.

РОЗДІЛ 1.

АНАЛІЗ СТАНУ ПРЕДМЕТНОЇ ГАЛУЗІ ТА ЗАВДАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ

1.1. Аналіз стану споживання електроенергії будівлями та обґрунтування доцільності її прогнозування

Орієнтовно 40% усіх енергетичних ресурсів на території України споживаються будівлями, до яких належить громадські будівлі – школи, заклади культури, будівлі державних установ, дитячих садків, лікарень. При цьому обсяги енергозбереження у окремих будівлях можуть становити орієнтовно 50...60% [1]. При цьому основною особливістю старих індивідуальних будинків є висока енергоємність. Індивідуальні будинки мають системи створення мікроклімату та санітарно-гігієнічних умов, що є також енергоспоживаючими.

В Україні існує низка інструментарію, які забезпечують регулювання діяльності щодо споживання електроенергії будівлями, а також їх енергомодернізації. Зокрема, це стосується розпоряджень [11], постанов [9] та законів [10; 4; 3]. Ними передбачено запровадження енерго сервісних договорів, які узгоджують процеси фінансування та можливого заощадження енергетичних ресурсів. При цьому виникає низка задач, які потребують зниження обсягів споживання ПЕР та підвищення комфортних умов у будівлях.

Передбачається, що розрахункові енергетичні моделі будівель розробляються на підставі обов'язкового проведення енергетичного аудиту будівель. Це є основою для формування систем енергетичного менеджменту будівель. Важливою умовою є потреба посилення уваги щодо виконання регулювань та забезпеченості мікроклімату індивідуальними будинками.

Рівень споживання електроенергії індивідуальними будинками залежить від великої множини чинників. Окремі із них згруповано та схематично зображено вплив чинників на рівень споживання електроенергії (рис. 1.1).



Рис. 1.1. Вплив чинників на рівень споживання електроенергії

Важливою є оцінка заздалегідь виконаних енергетичних обстежень індивідуальних будинків. За отриманими результатами енергетичних обстежень зроблять висновки щодо індивідуальних будинків:

- термічний опір огорожувальних конструкцій у 4-6 разів нижче чинних нормативів;
- не відповідають чинним вимогам системи опалення зокрема стосовно їх енергоефективності;
- зруйновані орієнтовно на 70% системи вентиляції, що не здатні забезпечити потрібний повітрообмін;
- у багатьох індивідуальних будинках не функціонує гаряче водопостачання, а також функціонує спрощено із використанням електро бойлерів.

За результатами проведених енергетичних аудитів індивідуальних будинків та інших будівель, виконавши проведений аналіз фактичного використання енергії будівлями, подано питомих фактичне споживання енергії рис. 1.2.

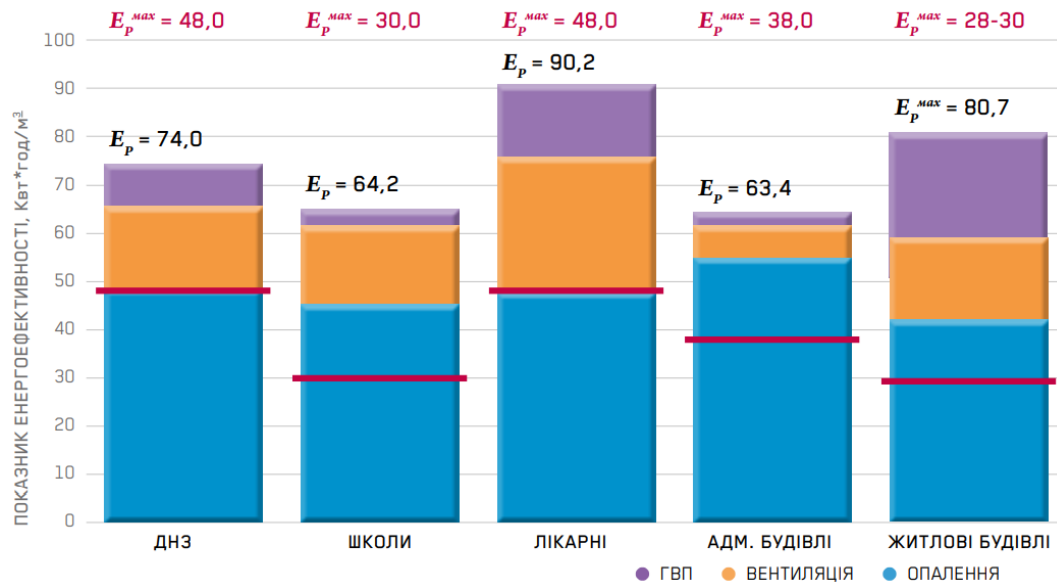


Рис. 1.2. Структура показників енергоефективності за результатами проведених енергоаудитів [1]

Для створення прийнятних і оптимальних умов перебування людей у індивідуальних будинках необхідно витратити більше енергії. Тому питання термомодернізації окремих будівель є до складними. Тому що ми маємо вирішити задачі зменшення енергоспоживання і водночас підвищення рівня комфорту. Підвищення повітрообміну в приміщенні до рівня допустимих і оптимальних параметрів без використання спеціальних вентиляційних систем з рекуперацією тепла, як показано на графіку на рисунку 2.3, призводить до зростання енергоспоживання.

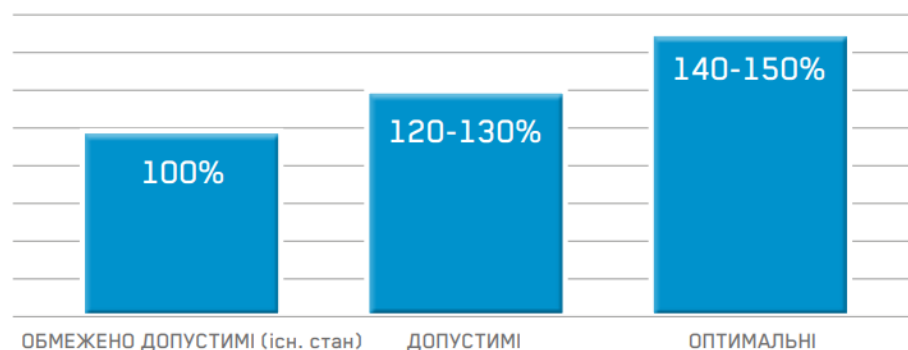


Рис. 1.3. Зростання енергоспоживання за підвищення повітрообміну в приміщеннях до рівня допустимих і оптимальних параметрів без використання спеціальних вентиляційних систем

Аналіз поданих вище результатів вказує на те, що фактична вартість електричної енергії пов'язана з поганим повітрообміном у приміщеннях будівель та відсутністю гарячого водопостачання, або не враховано витрати енергії на гаряче водопостачання. За попередньою оцінкою, загальне зниження споживання електроенергії енергії для шкільних будівель – 10%, дитячих садків – 12%, індивідуальних будівель – 10%.

На даний час існує задача точного прогнозування споживання електричної енергії для індивідуальних будівель, яка має важливе значення для вирішення низки прикладних задач, що пов'язані із процесами планування обсягів генерування та споживання електроенергії. Похибка такого прогнозу є причиною необґрунтованих витрат електроенергії індивідуальними будівлями. Це стосується того, що перебільшений прогнозований обсяг споживання електроенергії призводить до небажаної перевитрати енергії різними видами споживачів, а водночас зменшений прогнозований обсяг споживання електроенергії до зниження якості постачання енергії споживачами. Отже, точність прогнозування споживання електричної енергії є актуальною задачею сьогодення, яка водночас потребує використання ефективних підходів до її вирішення.

1.2. Методи прогнозування обсягів генерування та споживання електроенергії

Проведений аналіз відомих досліджень, а також публікацій, що стосуються прогнозування обсягів генерування та споживання електроенергії, свідчить про активність дослідників у зазначеному напрямі. Для цього використовуються різні підходи та методи, у основу яких покладено аналіз ретроспективних динамік генерування та споживання електроенергії, а також врахування різних чинників впливу на їх обсягів. Це забезпечує виявлення

статистичних зв'язків між чинниками для побудови прогнозних моделей. Окремі дослідними використовували різноманітні методи та програмні засоби.

Для прогнозування обсягів генерування та споживання електроенергії у 2000-х рр. розроблені програмні реалізації, які базуються на методах статистичного аналізу, виконанні прогнозів часових послідовностей (CatePillar, ForExSale, NCSS97 тощо). Усі вони належать до універсальних і не мають спеціалізації для конкретної предметної області. Це свідчить про те, що вони не можуть використовуватися диспетчерами як дорадчі інтегровані автоматизовані системи прогнозування обсягів генерування та споживання електроенергії.

З-поміж існуючих багатьох підходів, які досить часто використовуються для вирішення задачі прогнозування обсягів генерування та споживання електроенергії, можна зазначити традиційні (аналіз часових рядів – підходи Боксу-Дженкінса), а також використання регресійних алгоритмів, які вважаються ефективними методами обчислювального інтелекту (до них належать штучні нейронні мережі, тощо).

Доцільність використання обчислювального інтелекту в розглянутій задачі зумовлюється нелінійним характером прогнозованих обсягів генерування та споживання електроенергії, здатністю алгоритмів до самонавчання, а також виконання узагальнення, спостерігається високий рівень невизначеності (структурні та параметричні її види), наявністю стохастичності та хаотичності.

Сьогодні відомо низку наукових праць, які вміщують вдалі приклади використання нейронних мереж для вирішення задач прогнозування обсягів генерування та споживання електроенергії як в Україні, так і за її межами. В основному для цього прогнозуюча система базується на багат шаровому персептроні, який може мати різні модифікації, об'єднання із загальною архітектурою щодо передачі інформації.

Альтернативним підходом до прогнозування обсягів генерування та споживання електроенергії із використанням нейронних мереж із прямим розповсюдженням інформації є рекурентні нейронні мережі, які включають у

свою архітектуру глобальні та локальні (на рівні окремих шарів) зворотні зв'язки, а також передбачають навчання за допомогою реалізації спеціалізованих процедур [Помилка! Джерело посилання не знайдено.]. Вони забезпечують універсальні апроксимуючі і екстраполюючі можливості та здатні до навчання в умовах, що характеризуються суттєвою структурною і параметричною невизначеністю прогнозованих процесів споживання та генерування енергії.

Також наявні рекурентні нейронні мережі, які із обчислювального погляду є більш ефективні, порівняно із мережами прямої передачі інформації [Помилка! Джерело посилання не знайдено.]. Задачі обробки нелінійних часових рядів вирішувалися із використанням трьох типів рекурентних нейронних мереж: мережі Вільямса-Зіпсера [35], Елмана [19] і Джордана [Помилка! Джерело посилання не знайдено.].

Для вирішення задач, які стосуються аналізу та прогнозування часових рядів, що забезпечує виявлення змін властивостей складових цих рядів, запропоновані мережі вимагають модифікації. Насамперед це відноситься до алгоритмів навчання, так як означені нейронні мережі виконують навчання у пакетних режимах. При цьому не передбачено врахування того, що дані спостережень стосовно окремого часового ряду передбачають послідовне оброблення один за одним.

Стосовно прогнозування обсягів генерування та споживання електроенергії, то виконують різні типи прогнозування, які у переважній більшості різняться проміжком виконаного прогнозу. У нашій роботі розглядається короткострокове прогнозування обсягів генерування та споживання електроенергії.

1.3. Класифікація методів машинного навчання

Для прогнозування споживання обсягів генерування та споживання електроенергії застосовують різні методи. Переважно вони базуються на аналізі історичних даних щодо електроспоживання із врахуванням чинників, що на нього впливають. Це дає можливість виявити статистичні зв'язки між окремими ознаками даних та на їх основі виконати побудову прогнозних моделей.

Донедавна найбільш поширеними методами прогнозування у різних предметних галузях були однофакторні прогнози, які базуються на прогнозуванні часових рядів, що забезпечують використання регресійних методів. Однак, використання цих методів не дає можливість виконати точні прогнози через неможливість врахування впливу множини чинників на обсяги генерування та споживання електроенергії. Виникає потреба у практичному застосуванні багатфакторного прогнозування. Це виконання прогнозів забезпечує більшу точність порівняно із часовими рядами [7].

На рисунку 1.4 представлена класифікація методів машинного навчання за ознакою – тип задач.



Рис. 1.4. Класифікація методів машинного навчання

За типом вирішуваних задач (рис.1.4) методи машинного навчання поділяються на:

- навчання із учителем;
- навчання без учителя;
- навчання із частковим залученням учителя;
- навчання із підкріпленням.

Навчання із учителем. Являє собою контрольоване навчання, що є парадигмою машинного навчання для отримання інформації про взаємозв'язок вводу-виводу даних у систему на основі заданого набору парних навчальних вибірок вводу-виводу. Оскільки вихідний результат розглядається як мітка вхідних даних або спостереження, навчальна вибірка введення-виведення також називається позначеними навчальними даними або контрольованими даними. Іноді його також називають «Навчання з учителем», «Навчання із позначеними даними» або «Індуктивне машинне навчання».

Метою контрольованого навчання є побудувати штучну систему, яка може вивчати зв'язки між даними на вході і виході, що забезпечує прогнозування на виході системи з урахуванням нових входів. Якщо вихід приймає кінцевий набір дискретних значень, які вказують на мітки класу вхідних даних, навчене відображення призводить до класифікації вхідних даних. Якщо вихід приймає безперервні значення, це призводить до регресії входу. Інформація про взаємозв'язок «вхід-вихід» часто представлена параметрами моделі навчання. Коли ці параметри недоступні безпосередньо для навчання за вибірками системи, то слід пройти процес оцінки, щоб отримати бажані параметри.

На відміну від неконтрольованого навчання, навчальні дані для контрольованого навчання потребують контрольованої або позначеної інформації, тоді як навчальні дані для неконтрольованого навчання є неконтрольованими, оскільки вони не позначені (тобто це лише вхідні дані).

Якщо в алгоритмі використовуються навчальні дані як під контролем, так і над контролем, він називається алгоритмом напівконтрольованого навчання. Якщо алгоритм активно запитує користувача про мітки в процесі навчання, ітераційне контрольоване навчання, яке ще називають активним навчанням.



Рис. 1.5. Блок-схема навчання із учителем

На рисунку 1.5 показана блок-схема, яка ілюструє навчання із учителем. На цій діаграмі (x_i, y_i) – вибірка контрольованого навчання, де « x » представляє вхідні дані системи, « y » – вихідні дані системи (тобто контроль або маркування вхідних даних x), а « i » – індекс навчального зразка. Під час процесу навчання із учителем навчальний вхід x_i подається в систему навчання, а система навчання генерує вихід \tilde{y}_i . Вихід системи навчання \tilde{y}_i після цього порівнюється з основним істинним маркуванням y_i учителем, який обчислює різницю між ними. Різниця, яка на цій діаграмі називається сигналом помилки, потім надсилається до системи навчання для здійснюється коригування параметрів навчання. Метою цього процесу навчання є отримання набору оптимальних параметрів системи навчання, які можуть мінімізувати відмінності між \tilde{y}_i та y_i для всіх i -х значень, тобто мінімізувати загальну помилку для всього набору навчальних даних.

1.4. Оцінка енергетичної ефективності будівель за допомогою засобів машинного навчання

Останнім часом було проведено значний обсяг досліджень на тему енергетичної ефективності будівель (ЕЕБ). При цьому окремі із дослідників піднімали питання щодо зростання марнотратного використання енергії та її багаторічного негативного впливу на навколишнє середовище [21; 37].

Інструменти моделювання в даний час широко використовуються для аналізу або прогнозування енергоспоживання будівлі, щоб полегшити проектування та експлуатації енергоефективних будівель, оскільки практика показала, що результати моделювання часто можуть бути точними і відображають фактичні вимірювання [36].

Одними з них, які широко застосовувані у будівлях, є п'єзодатчики [23]. Вони використовуються у індивідуальних будинках для зменшення споживання енергії [18].

Програмне забезпечення може надати надійні рішення для оцінки впливу альтернатив проектування будівель, однак цей процес може бути дуже трудомістким і потребуватиме досвіду користувача в конкретній програмі. Отже, на практиці багато дослідників покладаються на інструменти машинного навчання для вивчення впливу різних побудова параметрів (наприклад, компактність) на деяких змінних, що цікавлять (наприклад, енергія), оскільки це простіше та швидше, якщо доступна база даних необхідних діапазонів змінних [29]. Різні методи машинного навчання, такі як поліноміальна регресія [16], опорний вектор машини (SVM) [17], штучні нейронні мережі (ANN) [38] і дерева рішень [37] були досліджені для прогнозування різних величин, що представляють інтерес у контексті ЕЕБ. Також вони використовувалися для прогнозування HL і CL [29].

Методи машинного навчання (HL і CL) для окремих індивідуальних будівель потребують окремого дослідження. У цьому випадку класична лінійна регресія як базовий метод і дві різні моделі класифікації та регресії розглядаються як базові. У дослідженнях [29] обґрунтовано дві моделі логістичної регресії та випадкового лісу.

На підставі вище сказано встановлено, що слід провести дослідження методів машинного навчання, які забезпечують прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками на підставі історичних даних. Це забезпечить підвищення ефективності прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками завдяки обґрунтуванню раціональної моделі, що базується на алгоритмі машинного навчання.

1.5. Завдання кваліфікаційної роботи

На підставі сказаного, можна вважати, що використати існуючий інструментарій прогнозування часових рядів неможна, вони мають недоліки, що знижує точність виконаного прогнозу. Це зумовлює потребу дослідження моделей машинного навчання для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками. Для цього слід вирішити такі завдання:

- виконати аналіз існуючих методів машинного навчання та сформулювати завдання кваліфікаційної роботи;
- вибрати для дослідження моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками із використанням методів машинного навчання;
- обґрунтувати ефективну модель для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками;
- виконати прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками із використанням методів машинного навчання;
- розробити заходи із охорони праці та безпеки у надзвичайних ситуаціях;
- виконати розрахунок економічної ефективності від використання

запропонованої моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.

РОЗДІЛ 2.

ВИБІР МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ОБСЯГІВ ГЕНЕРУВАННЯ ТА СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ ІНДИВІДУАЛЬНИМИ БУДИНКАМИ

2.1. Метод дерев рішень (Decision Tree Regressor)

Алгоритм дерева рішень (DT) є потужним і універсальним, оскільки він дуже ефективний для фіксації взаємодії між функціями та його легко візуалізувати, що робить їх ефективними для класифікації та прогнозування даних для вирішення різних прикладних задач [1].

Дерево рішень складається з кінцевого вузла та внутрішнього вузла (вузла прийняття рішень), включаючи кореневий вузол. DT використовує алгоритм «Розділяй і володарюй». По-перше, DT ініціалізує кореневий вузол, який потім розбивається на дві підмножини, обчислені на основі критерій розбиття, такий як приріст інформації, індекс Джині для завдання класифікації або середня квадратична помилка (MSE), Пуассона для завдання регресії. Розподіли класифікуються як термінальний (кінцевий вузол) або рішення (внутрішній вузол) на основі критерію зупинки, яким є кожен внутрішній вузол обробляються окремо.

Коли дерево побудовано на основі вхідних даних, воно використовується для завдань прогнозування та класифікації. В обох випадках модель проходить дерево, починаючи від кореневого вузла до кінцевого (кінцевого вузла). Точки даних є середніми залежними змінними для цього кінцевого вузла [27].

Як показано на рис. 2.1, Decision Tree Regressor прогнозує середні значення для всіх точок даних у сегменті (де кожен сегмент представляє листовий вузол). На рис. 2.1 ми можемо легко підігнати лінійну лінію з лівого боку, забезпечуючи краще представлення даних.

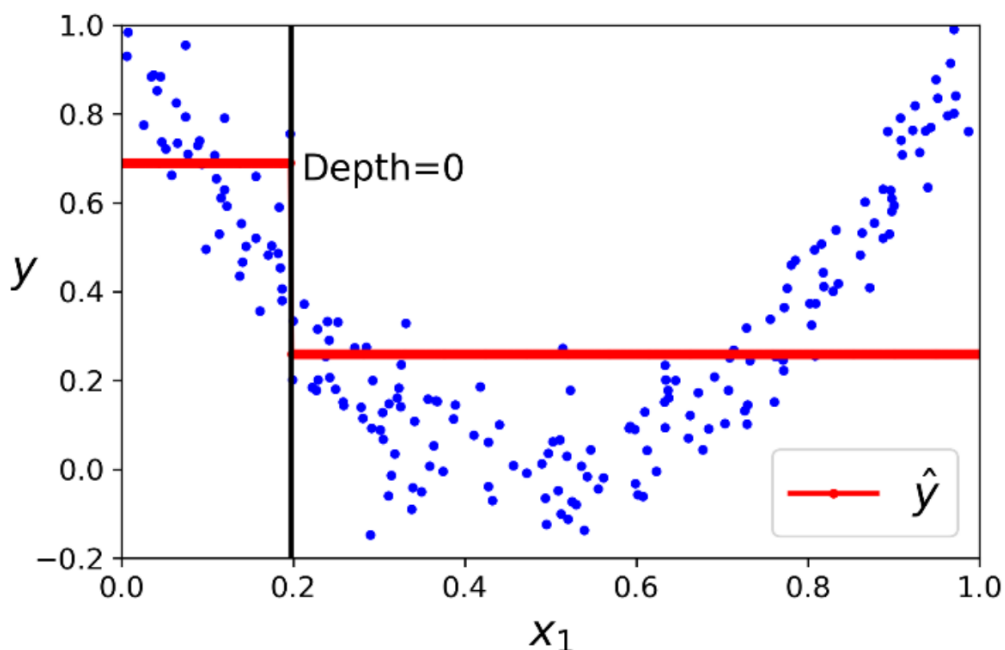


Рис. 2.1. Прогнозування із використанням дерева рішень

Регресор лінійного дерева рішень (запропонована модель) (LDTR) не тільки підбирає лінійну лінію (лінійна регресія) за наявними даними, але також використовує модифікований алгоритм поділу вузлів, щоб знайти розбиття таким чином, щоб кожен вузол найкраще фіксував лінійне співвідношення в кожному сегменті. На рис. 2.2 показано, як LDTR будує дерево на тому ж наборі даних, що використовується на рис. 2.1.

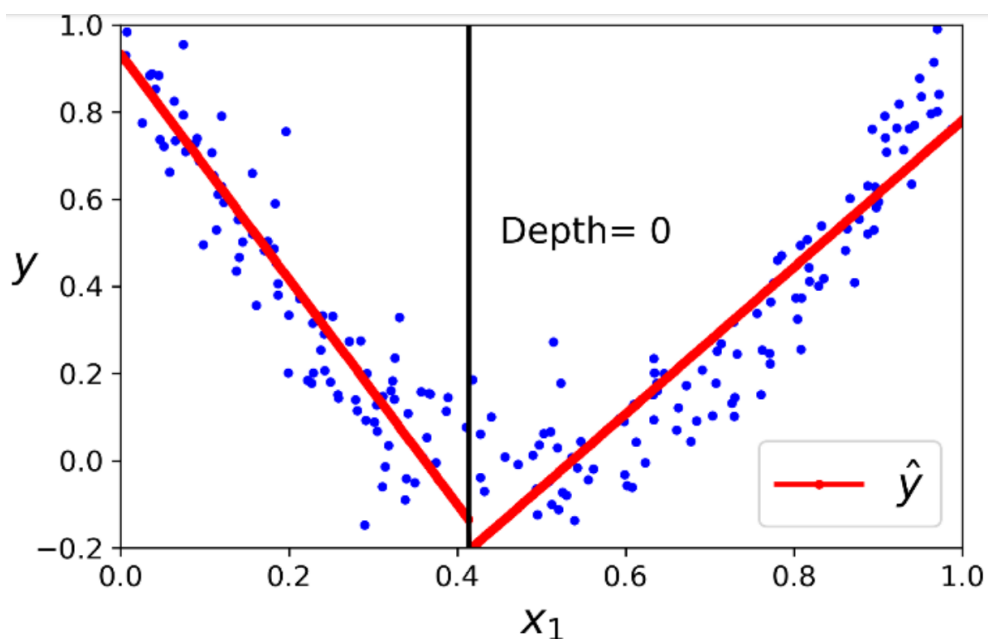


Рис. 2.2. Прогноз регресії лінійного дерева рішень

Дерево рішень

Decision Tree Regressor побудовано на алгоритмі CART, запропонованому Брейманом та іншими в 1984 р. [15], програма, яка використовується в моделі Decision Tree.

Два основних компоненти, які будуть належать до дерев рішень:

1) Структура вузла:

Кожен вузол зберігає атрибути, які визначають структуру дерева та допомагають у обході дерева під час прогнозування, наприклад:

- I. left_child: ідентифікатор лівого дочірнього елемента вузла.
- II. right_child: ідентифікатор правого дочірнього вузла,
- III. feature: функція, яка використовується для розбиття вузла.
- IV. threshold: Порогове значення у вузлі.

2) Розділення вузла:

Функція Node split знаходить функцію та порогове значення для поділу вузла. Вона обчислює функцію вартості CART для різних функцій k і порогового значення t_k , а потім розділяє вузол для функції та порогового значення, яке мінімізує функцію вартості.

Функція вартості CART для регресії має вигляд:

$$J(k, t_k) = \frac{m_{left}}{m} MSE_{left} + \frac{m_{right}}{m} MSE_{right}, \quad (2.1)$$

$$were \begin{cases} MSE_{node} = \sum_{i \in node} (\hat{y}_{node} - y^i)^2 \\ \hat{y}_{node} = \frac{1}{m_{node}} \sum_{i \in node} y^i \end{cases} \cdot \quad (2.2)$$

Коли метод поділу вузла знаходить функцію та поріг, які мінімізують MSE, він зберігає ці значення у вузлі.

На рис. 2.3 представлена блок-схема запропонованої моделі.

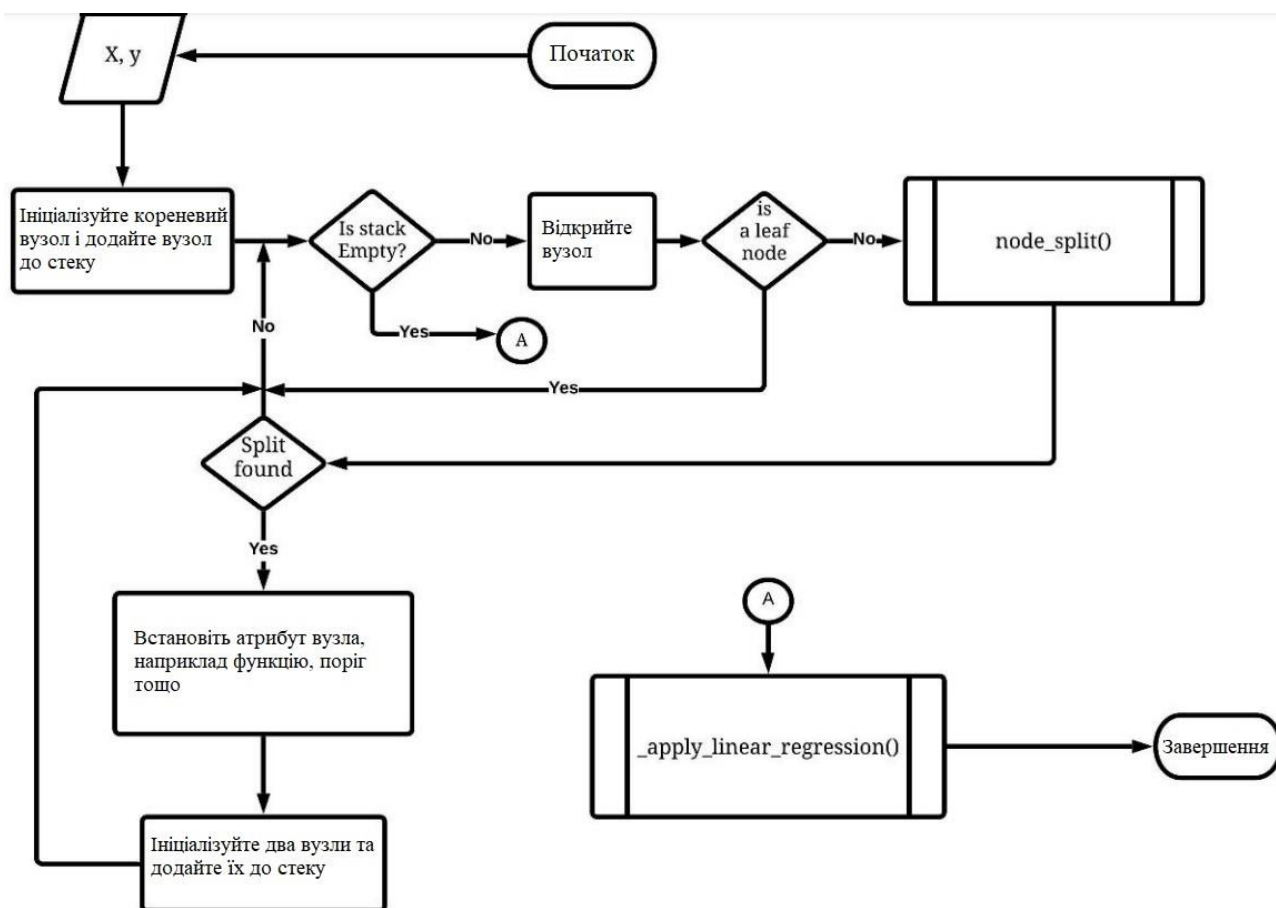


Рис. 2.3. Блок-схема алгоритму LDTR

Однак його можна використовувати для розуміння потоку для існуючої регресорної моделі дерева рішень із кількома змінними. Поточна модель не має методу `_apply_linear_regressor()` і використовує інший метод `node_split()`.

Лінійна регресія

Для оцінки зв'язку між залежними та незалежними змінними використовується лінійна регресія [14]. Вона знаходить коефіцієнти:

$$y = wX + b, \quad (2.3)$$

$$w = \text{coefficient } b = \text{intercept}$$

$$w = (w_1, \dots, w_n), \quad (2.4)$$

Коефіцієнти (2.4) найкраще представляють лінійну залежність. Зокрема, передбачають залежне значення шляхом множення цих коефіцієнтів на відповідні незалежні змінні, як показано в рівняннях 2.2 і 2.3 [22].

Множинна лінійна регресія має вигляд:

$$y = \sum_{j=1}^n w_j X_j + b,$$

$$w_j = \text{coefficient for variable } X_j ,$$

$$b = \text{intercept}$$
(2.5)

Однак нелінійну лінію можна точно представити як набір кінцевих (нескінченних лінійних) ліній. LDTR розбиває нелінійну лінію на набір лінійних сегментів. Як показано на рис. 2.2, дві незалежні лінійні лінії точно представляють нелінійну лінію. Додавання більше ліній підвищить точність.

На рис. 2.4 показано набір даних, який повторює синусоїдальну криву над $[-\pi, \pi]$.

1.5

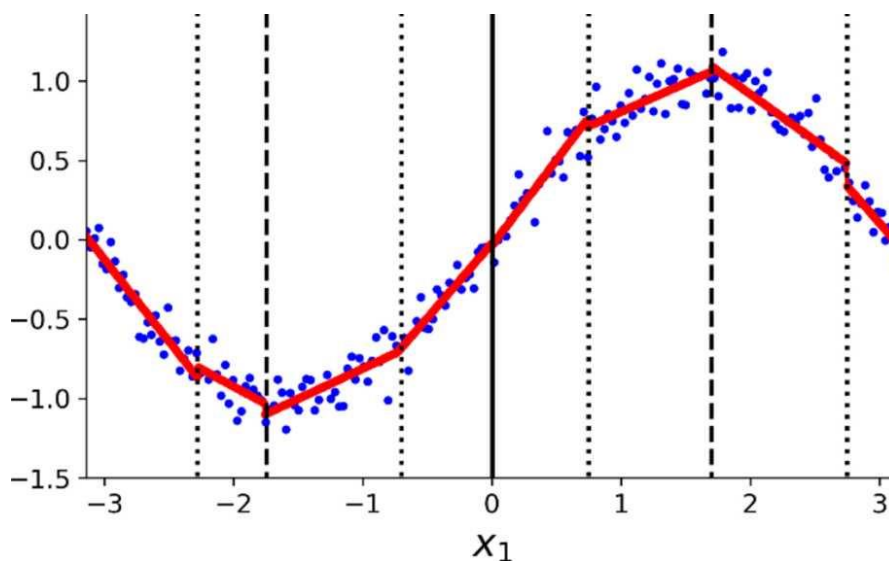


Рис. 2.4. Прогнозування із використанням LDTR на синусоїдальній кривій

Одна лінійна лінія не може відповідати набору даних. Однак, як показано на рис. 2.4, набір із восьми чітких прямих ліній може представляти криву з високою точністю. LDTR можна використовувати для обчислення набору прямих ліній для представлення нелінійних даних. LDTR забезпечує розбиття та опис лінійних регресій для кожного сегмента, показано на рис. 2.4.

Подібно до лінійної регресії [26], LDTR також чутливий до викидів. LDTR також має вищу часову складність порівняно з поліноміальною

регресією (лінійна регресія з поліноміальною ознакою). Однак LDTR, на відміну від поліноміальної регресії, не потребує додаткового аналізу чи підготовки, щоб знайти ступінь (розмір), у якому дані є лінійними, а потім трансформувати дані до цього ступеня. Отже, хоча поліноміальна регресія має меншу часову складність, LDTR все одно може бути наближеною альтернативою.

2.2. Регресія випадкового лісу (Random Forest Regressor)

Дерева рішень дуже добре підходять для опису нелінійних зв'язків між вхідними функціями та цільовою змінною. Внутрішню роботу дерева рішень можна розглядати як групу умов if-else.

Починається аналіз із самого верху одним вузлом (рис. 2.5).

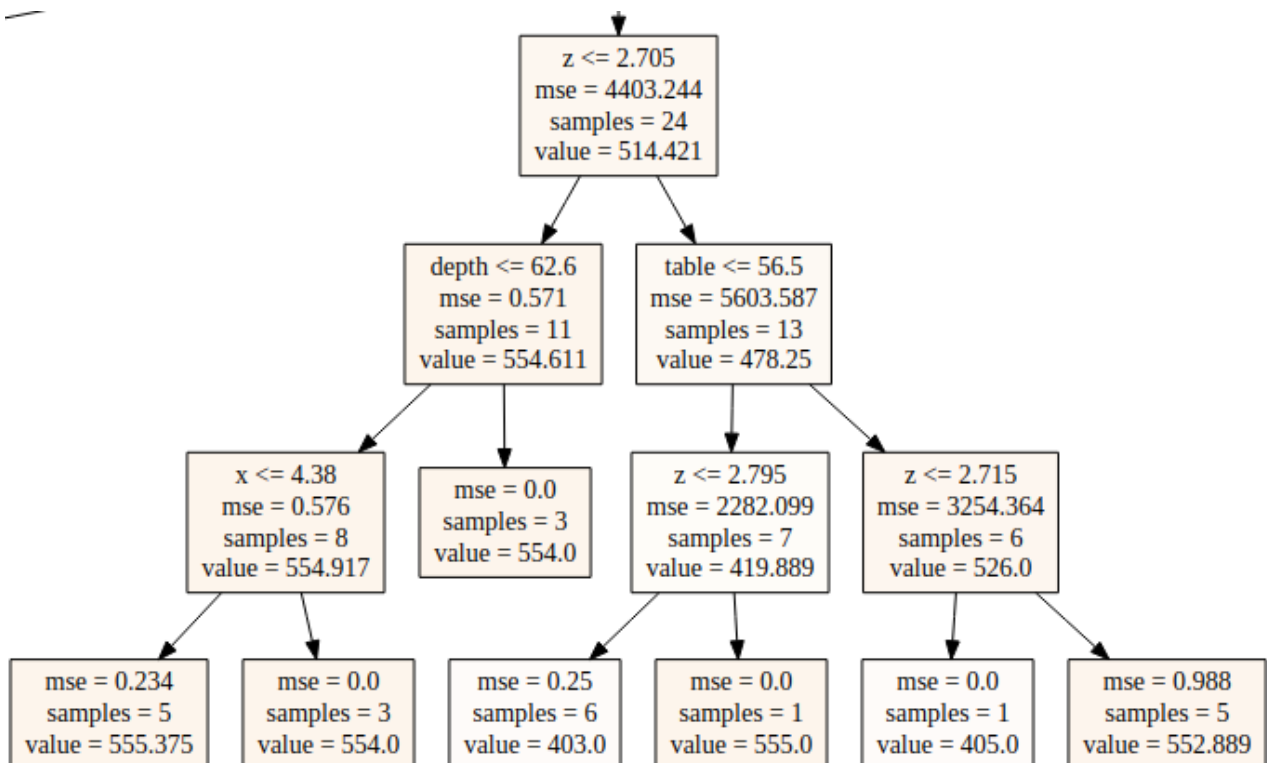


Рис. 2.5. Приклад регресії дерева рішень

Потім цей вузол розбивається на лівий і правий окремі вузли – вузли прийняття рішень. У подальшому ці вузли розділяються на відповідні лівий та правий вузли, які є складовими попереднього. У кінці листового вузла обчислюється середнє значення спостереження, яке відбувається із прив'язкою до контрактного об'єкту. Найнижчі вузли називаються листками або кінцевими вузлами. Значення у листках належить до середніх значень спостережень, що відбуваються у заданому об'єкті.

Наскільки глибоко розглядається розщеплення дерева, така і є глибина дерева. Це один із гіперпараметрів, який можна налаштувати. Максимальна глибина дерева вказана, щоб запобігти надто глибокому дереву – сценарій, який призводить до перенавчання.

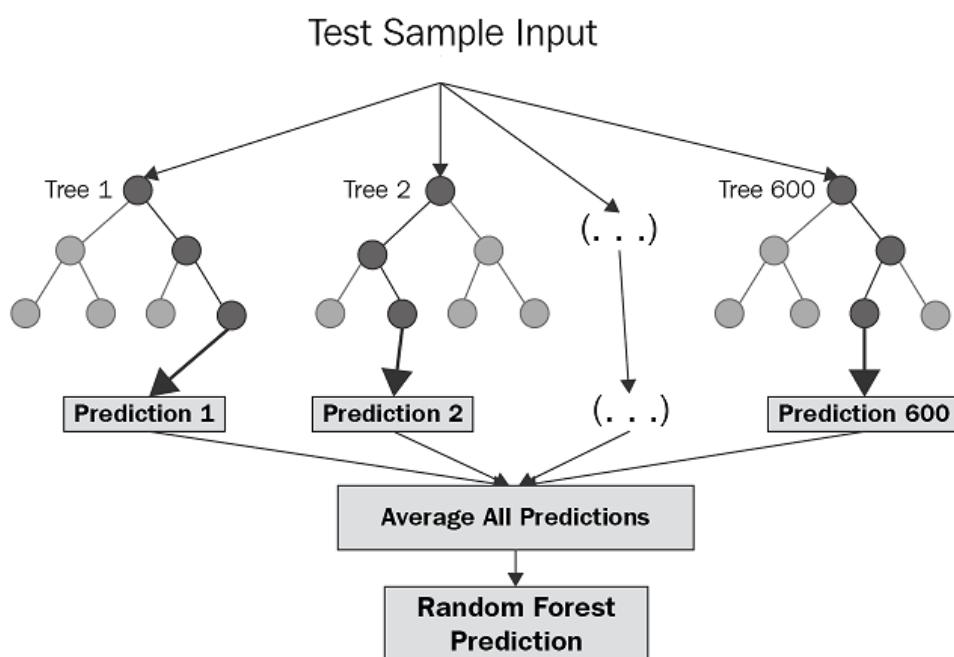


Рис. 2.6. Схема регресії випадкового лісу

Регресія випадкового лісу

Випадковий ліс являє ансамбль дерев рішень. Це означає, що багато дерев, побудованих певним «випадковим» чином, утворюють випадковий ліс. Кожне дерево створюється з різної вибірки рядків, і в кожному вузлі для

розділення вибирається різна вибірка функцій. Кожне з дерев робить свій індивідуальний прогноз. Потім ці прогнози усереднюються для отримання єдиного результату (рис. 2.6).

Усереднення робить випадковий ліс кращим за окреме дерево рішень, отже, покращує його точність і зменшує перенавчання. Прогноз Random Forest Regressor – це середнє значення прогнозів, зроблених деревами в лісі.

2.3. Багат шаровий перцептрон (MLP Regressor)

Одним із показників, яким багат шаровий перцептрон MLP відрізняється від інших методів тим, що існує багато параметрів, які можна налаштувати та згрупувати за двома критеріями. По-перше, щоб виконати різні дослідження, а по-друге, щоб отримати кращу продуктивність та результат прогнозування. Розглянемо коротко структуру MLP (рис. 2.7).

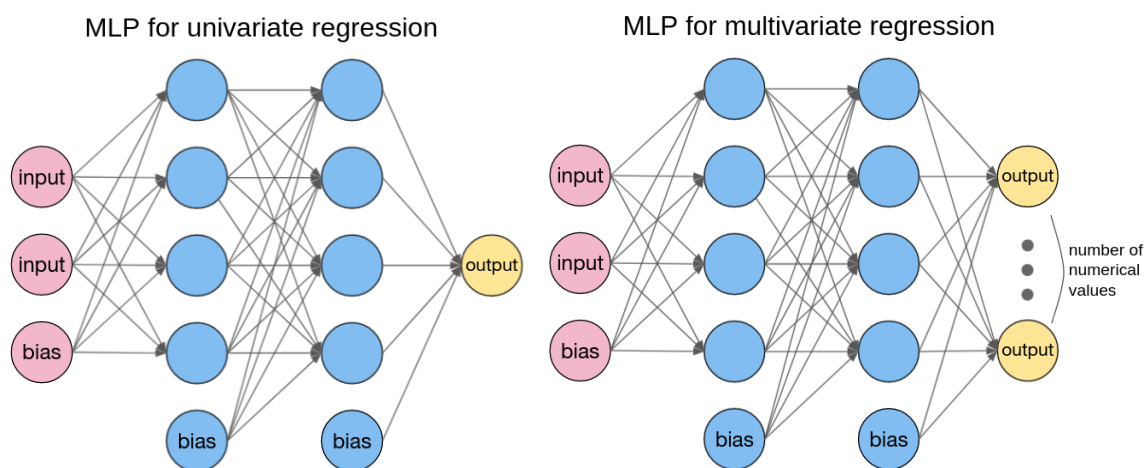


Рис. 2.7. Структура багат шарового перцептроні MLP

Залежно від форми вихідного шару, є одно та багато структурний. Переважаючим є один вихідний нейрон на кожне значення, яке передбачено MLP. Тому для однофакторної регресії вихідний рівень MLP потребує лише

одного вихідного нейрона. Щоб передбачити багато функціональну складову, MLP потребує двох і більше вихідних нейронів.

Функція активації вихідного нейрона залежить від того, який вихід ви хочете отримати. Коли використовується функція активації у вихідному нейроні, здійснюється обмеження діапазону виведення. Тому, як правило, не включається функція активації у вихідний нейрон. Однак бувають випадки, коли вони є потрібні. Наприклад, якщо виконується прогнозування вартості, не має сенсу мати від'ємний результат. У цьому випадку можна використовувати Relu або Softplus, щоб переконатися, що кожен результат є додатним.

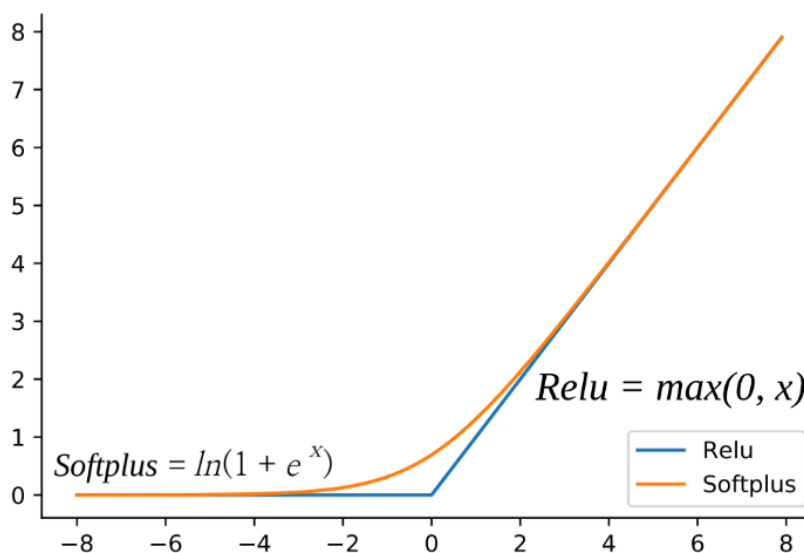


Рис. 2.8. Графіки функцій «Relu» та «Softplus»

З іншого боку, під час прогнозування ймовірності настання заданої події існує потреба у тому, щоб вихід був у діапазоні від 0 до 1 або від 0 до 100. У цьому випадку можна використовувати функції «Logistic» або «Tanh» і відповідно масштабувати, щоб отримати бажану межу показника на виході (рис 2.9).

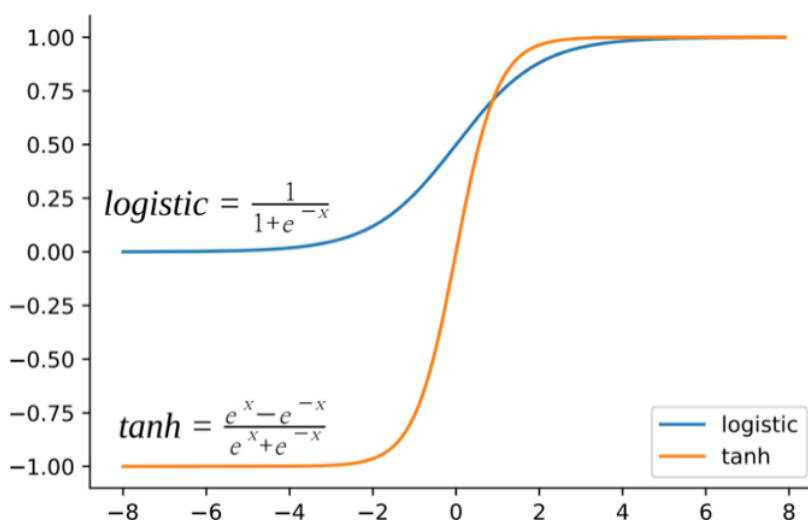


Рис. 2.9. Графіки функцій «Logistic» та «Tanh»

Загальна метрика – це середня квадратична помилка, яка обчислює середнє значення (фактично-прогнозованого у квадраті). Однак, якщо існує багато викидів, зведення різниці між фактичним і прогнозованим неминуче дасть високе значення, навіть якщо модель добре підходить за відсутності викидів. Тож у цьому випадку для оцінювання моделі можна використати середню абсолютну похибку:

$$MSE = \frac{1}{n} \sum (y - \hat{y})^2, \quad (2.6)$$

$$MAE = \frac{1}{n} \sum |y - \hat{y}|. \quad (2.7)$$

2.4. Підвищення градієнта (Gradient Boosting Regressor)

Підвищення градієнта є одним із варіантів методів ансамблю, коли ви створюєте кілька слабких моделей і комбінуйте їх, щоб отримати кращу продуктивність у цілому. Для цього слід створити дерева регресії досліджуваного об'єкта з посиленням градієнта, який має нелінійний зв'язок між x та y (рис. 2.10).

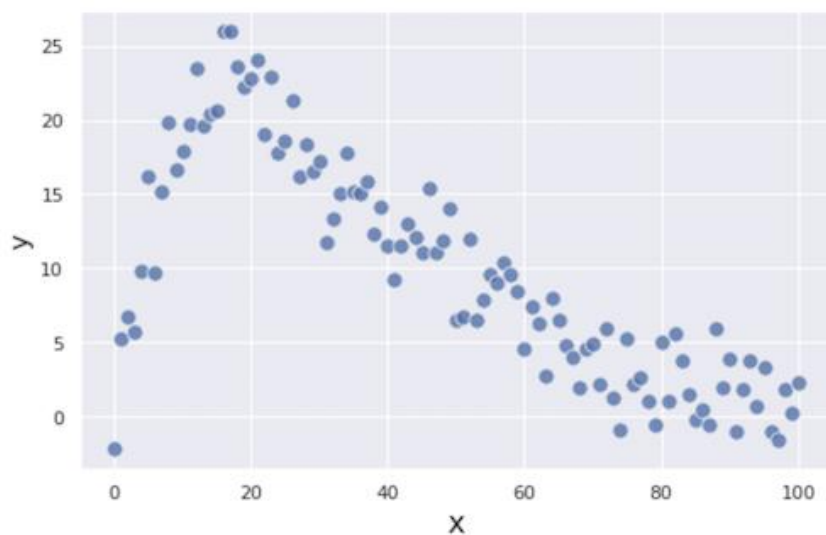


Рис. 2.10. Приклад досліджуваної функції

На першому етапі слід зробити дуже наближене передбачення цілі y . Ми робимо початковий прогноз F_0 як загальне середнє значення y :

$$F_0 = \text{mean}(y). \quad (2.8)$$

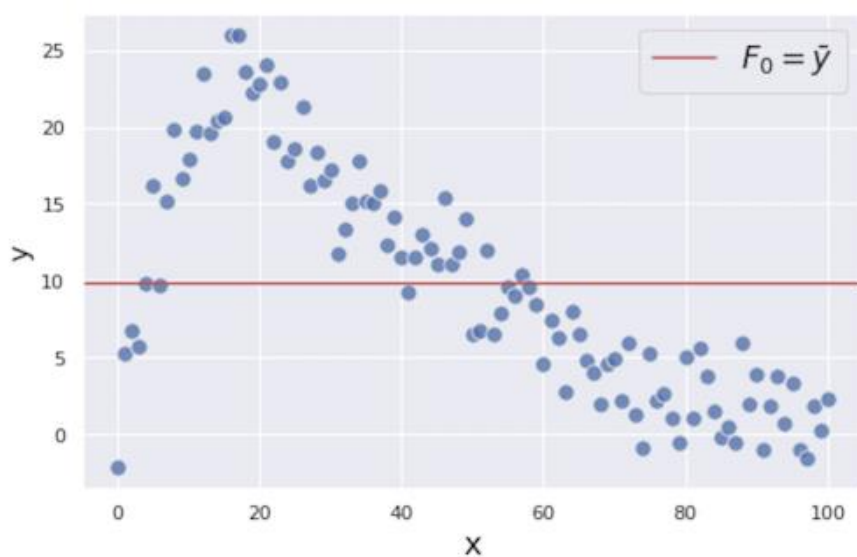


Рис. 2.11. Результати виконання початкового прогнозу

Можна вважати, що використовувати середнє значення $\text{mean}(y)$ для прогнозу є недостатньо точним. Ми покращимо наш прогноз, додавши до нього більше малих моделей (рис. 2.12).

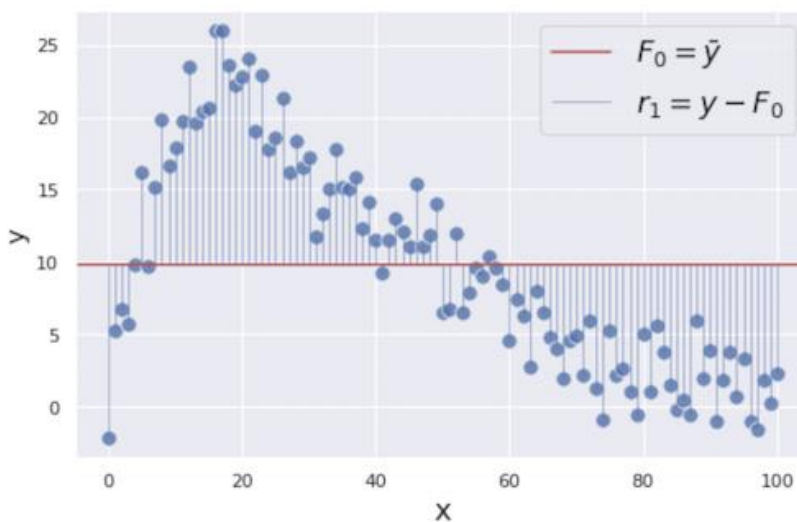


Рис 2.12. Аналіз відхилень у прогнозній моделі

Щоб покращити потрібний прогноз, ми зосередимося на залишках (тобто помилках прогнозування) із першого етапу, оскільки це те, що можна мінімізувати, щоб отримати кращий прогноз. Відхилення r_1 показані вертикальними синіми лініями на рис 2.12.

Щоб мінімізувати ці відхилення, будується модель регресійного дерева із x її ознакою та відхиленнями:

$$r_1 = y - \text{mean}(y). \quad (2.9)$$

Це пояснюється тим, що якщо ми зможемо знайти деякі закономірності між x та r_1 , побудувавши додаткову модель, ми зможемо зменшити відхилення, на підставі її використання.

Для візуалізації будуємо дуже прості дерева, кожне з яких має лише одне розвітлення і два кінцевих вузла. Деревя підсилення градієнта зазвичай мають трохи глибші дерева, такі як дерева з 8...32 кінцевими вузлами. Тут ми створюємо перше дерево, яке прогнозує відхилення з двома різними значеннями:

$$r_1 = \{6.0, -5.9\}. \quad (2.10)$$

У зазначеному прикладі використано γ (гама) для позначення прогнозу досліджуваного показника (рис. 2.13).

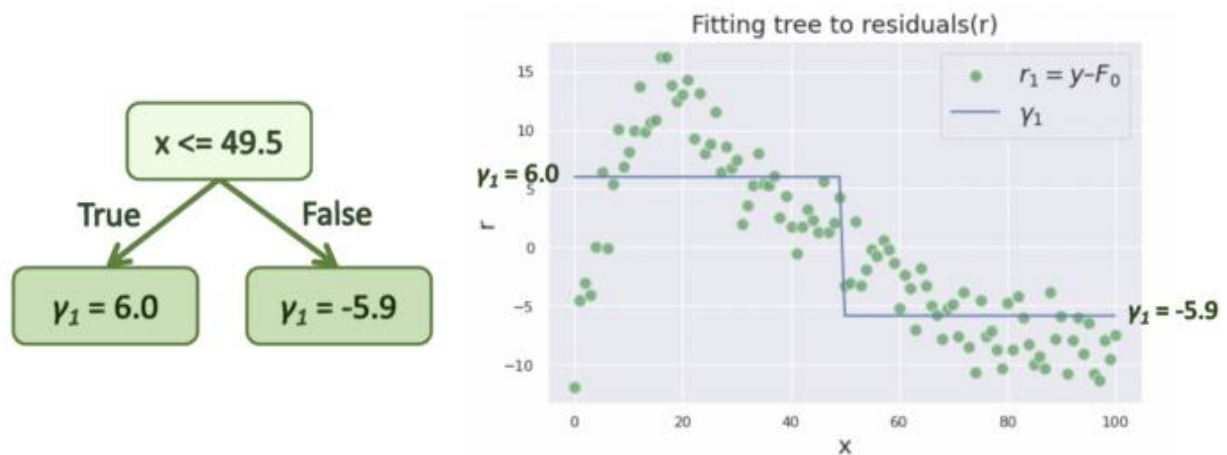


Рис. 2.13. Прогнозування відхилення з двома різними значеннями

Отриманий прогноз γ_1 додається до нашого початкового прогнозу F_0 , щоб зменшити відхилення від емпіричних даних. Насправді, алгоритм посилення градієнта не просто додає γ , а F_0 робить модель наближеною для навчальних даних. Натомість γ зменшується за швидкістю навчання, ν яка коливається у межах від 0 до 1, а потім додається до F_0 :

$$F_1 = F_0 + \nu \cdot \gamma_1. \quad (2.11)$$

У цьому прикладі ми використовуємо відносно велику швидкість навчання, $\nu = 0.9$ щоб полегшити розуміння процесу оптимізації, але зазвичай має бути набагато меншим значенням, наприклад 0,1.

Після оновлення наш комбінований прогноз F_1 матиме вигляд:

$$F_1 = \begin{cases} F_0 + \nu \cdot 6.0 & \text{if } x \leq 49.5 \\ F_0 - \nu \cdot 5.9 & \text{otherwise} \end{cases}. \quad (2.12)$$

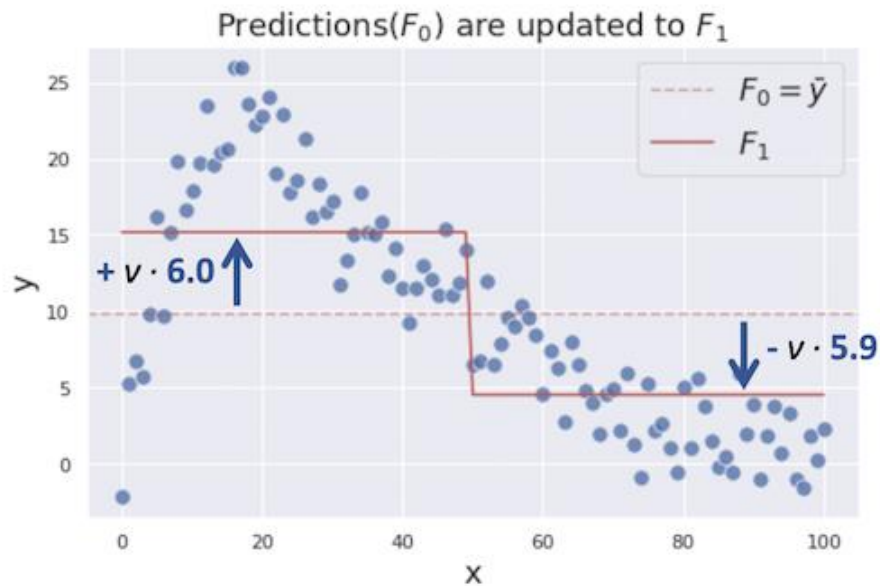


Рис. 2.14. Комбінований прогноз відхилення з двома різними значеннями

Тепер оновлені відхилення r_2 виглядають так, як показано на рис. 2.15.

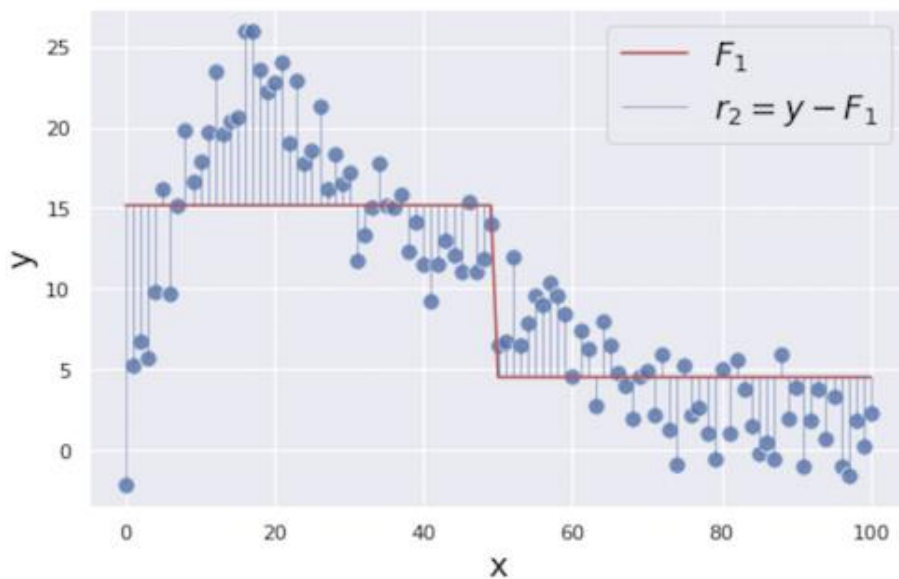


Рис. 2.15. Комбінований прогноз із оновленими відхиленнями

На наступному етапі знову створюється дерево регресії, використовуючи те саме x , що й функція, із оновленими відхиленнями r_2 . При цьому створено нове дерево (рис. 2.16).

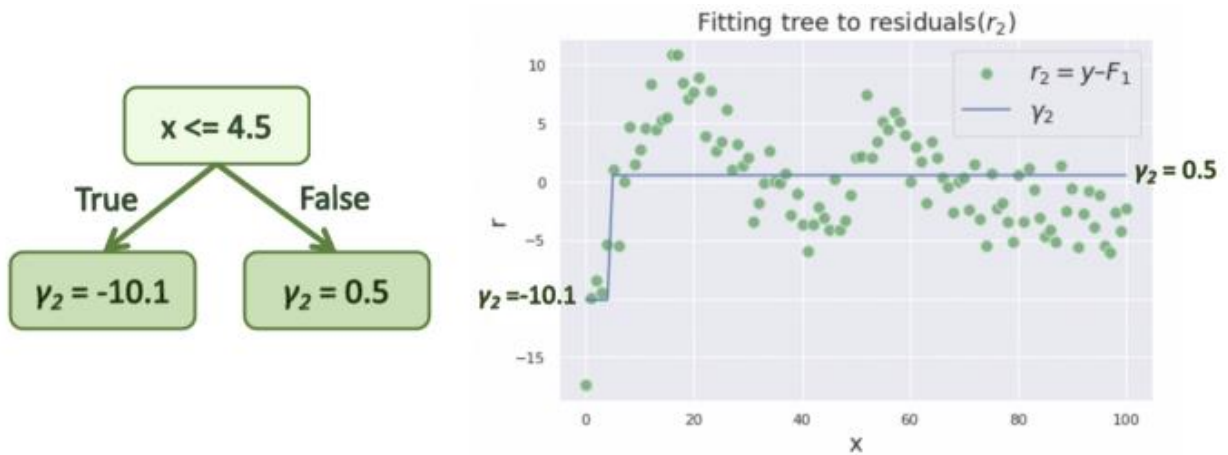


Рис. 2.16. Оновлене прогнозування відхилення з двома різними значеннями

Після цього виконуємо оновлення попередньо комбінованого прогнозування F_1 новим прогнозуванням дерева γ_2 (рис. 2.17).

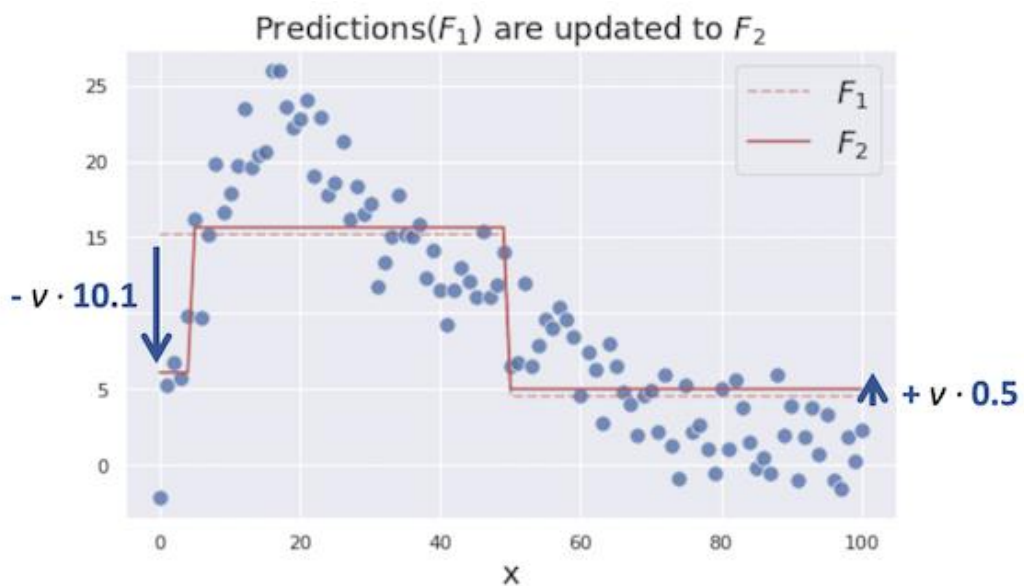


Рис. 2.17. Удосконалене прогнозне дерево

У подальшому зазначені етапи повторюються доти, доки прогнозована модель не перестане покращуватися. Нижче на рис. 2.18 показано процес оптимізації від 0 до 6 ітерацій.

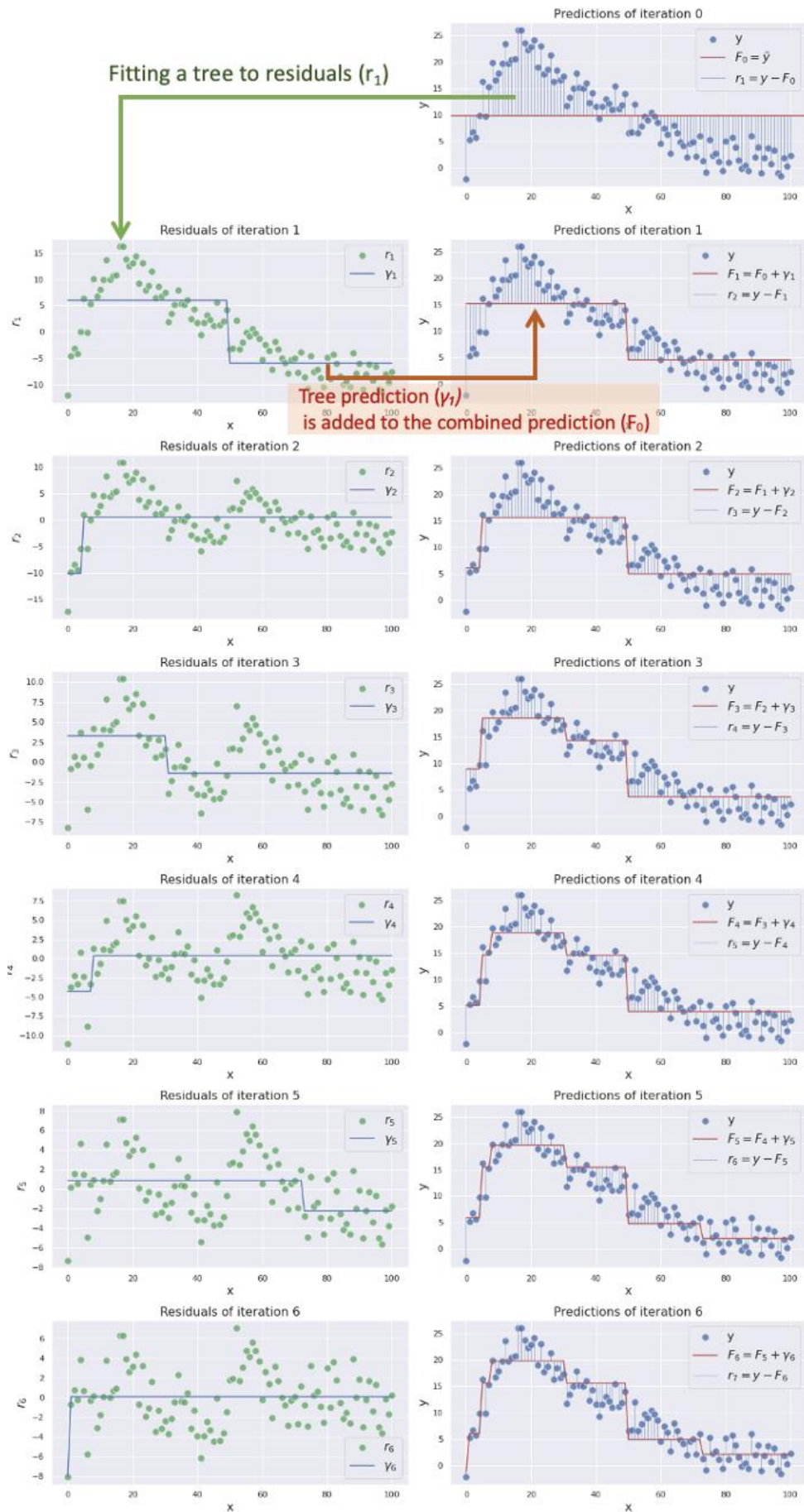


Рис. 2.18. Процес оптимізації моделі прогнозування (від 0 до 6 ітерацій)

У результаті отримується комбіноване прогнозування F_m досліджуваного показника, яке стає все ближчим до поставленої мети, оскільки додається більше дерев до комбінованої моделі. Означений працює градієнтного посилення для моделей прогнозування складних об'єктів шляхом поєднання кількох менших моделей.

РОЗДІЛ 3.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ОБСЯГІВ ГЕНЕРУВАННЯ ТА СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ ІНДИВІДУАЛЬНИМИ БУДИНКАМИ

3.1. Підготовка та аналіз даних для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками

Для виконання прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками обрано загальнодоступний набір даних [20]. У цих даних представлено показники для оцінювання вимог щодо навантаження на системи опалення та охолодження будівель (тобто енергоефективності) як функції параметрів енергоефективності індивідуальних будинків.

У аналізованому датасеті присутні колонки:

1. Відносна компактність – X1 Relative Compactness;
2. Площа поверхні, м² – X2 Surface Area;
3. Площа стін, м² – X3 Wall Area;
4. Площа даху, м² – X4 Roof Area;
5. Загальна висота, м – X5 Overall Height;
6. Орієнтація (2: Північ, 3: Схід, 4: Південь, 5: Захід) – X6 Orientation;
7. Площа скління (0%, 10%, 25%, 40% площі підлоги) – X7 Glazing Area;
8. Розподіл площі скління (відхилення) (1: Рівномірний, 2: Північ, 3: Схід, 4: Південь, 5: Захід) – X8 Glazing Area Distribution;
9. Навантаження на опалення, кВт.год/м² – Y1 Heating Load;
10. Навантаження на охолодження, кВт.год/м² – Y2 Cooling Load.

Представлений набір даних містить 8 вхідних змінних ($X_1 \dots X_8$) та 2 вихідних змінних (Y_1 та Y_2). Нас цікавить цей набір даних, щоб вивчити три цікаві напрямки щодо отримання знань із даних, а також їх візуалізувати та встановити як вони впливають на енергетичне навантаження індивідуальних будинків.

Для попереднього аналізу даних використовували графічний інтерфейс для швидкої роботи з наборами даних Mito, який за окремими діями створює код на Python. Спостерігаючи за кодом Mito, можна дізнатися щось нове, при цьому зрозумілий код із багатьох рядків можна отримати всього за кілька кліків.

Представлений набір даних із параметрами енергоефективності індивідуальних будинків опрацьовуємо із використанням окремого блокноту JupyterLab notebook. Насамперед імпортуємо бібліотеку рядком `import mitosheet`, і запускаємо її командою `mitosheet.sheet()` (рис. 3.1).

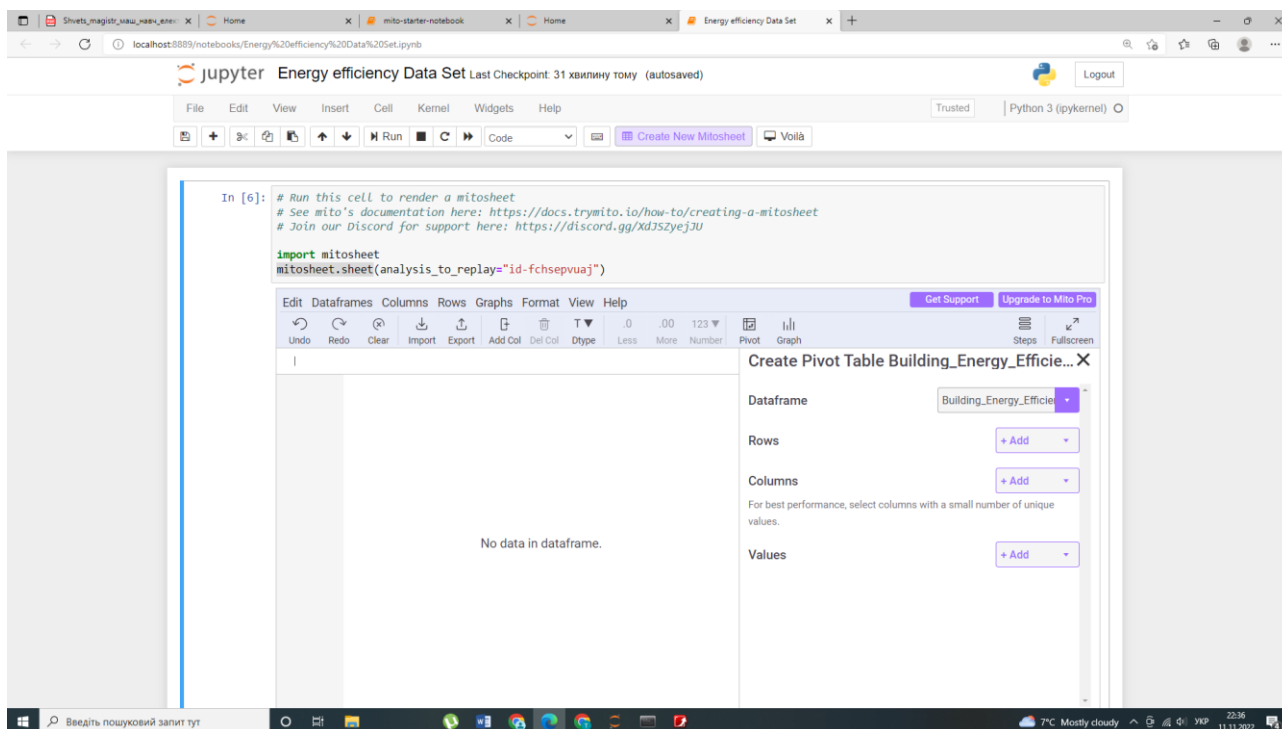


Рис. 3.1. Вікно графічного інтерфейсу Mito у блокноті JupyterLab notebook

Попередній аналіз даних свідчить про те, що площа стіни, площа даху, площа скління є ключовими показниками та можуть впливати на ефективність

енергетичного навантаження для обох енергетичних систем (опалення та охолодження) індивідуальних будинків (рис. 3.2).

	Relative Compactness	Surface Area	Wall Area	Roof Area	Overall Height	Orientation	Glazing Area	Glazing Area Distribution	Heating Load	Cooling Load
5	0.90	563.50	318.50	122.50	7.00	3	0.00	0	21.46	25.38
6	0.90	563.50	318.50	122.50	7.00	4	0.00	0	20.71	25.16
7	0.90	563.50	318.50	122.50	7.00	5	0.00	0	19.68	29.60
8	0.86	588.00	294.00	147.00	7.00	2	0.00	0	19.50	27.30
9	0.86	588.00	294.00	147.00	7.00	3	0.00	0	19.95	21.97
10	0.86	588.00	294.00	147.00	7.00	4	0.00	0	19.34	23.49
11	0.86	588.00	294.00	147.00	7.00	5	0.00	0	18.31	27.87
12	0.82	612.50	318.50	147.00	7.00	2	0.00	0	17.05	23.77
13	0.82	612.50	318.50	147.00	7.00	3	0.00	0	17.41	21.46
14	0.82	612.50	318.50	147.00	7.00	4	0.00	0	16.95	21.16
15	0.82	612.50	318.50	147.00	7.00	5	0.00	0	15.98	24.93
16	0.79	637.00	343.00	147.00	7.00	2	0.00	0	28.52	37.73
17	0.79	637.00	343.00	147.00	7.00	3	0.00	0	29.90	31.27
18	0.79	637.00	343.00	147.00	7.00	4	0.00	0	29.63	30.93
19	0.79	637.00	343.00	147.00	7.00	5	0.00	0	28.75	39.44
20	0.76	661.50	416.50	122.50	7.00	2	0.00	0	24.77	29.79
21	0.76	661.50	416.50	122.50	7.00	3	0.00	0	23.93	29.68
22	0.76	661.50	416.50	122.50	7.00	4	0.00	0	24.77	29.79
23	0.76	661.50	416.50	122.50	7.00	5	0.00	0	23.93	29.40
24	0.74	686.00	245.00	220.50	3.50	2	0.00	0	6.07	10.90
25	0.74	686.00	245.00	220.50	3.50	3	0.00	0	6.05	11.19
26	0.74	686.00	245.00	220.50	3.50	4	0.00	0	6.01	10.94
27	0.74	686.00	245.00	220.50	3.50	5	0.00	0	6.04	11.17
28	0.71	710.50	269.50	220.50	3.50	2	0.00	0	6.37	11.27
29	0.71	710.50	269.50	220.50	3.50	3	0.00	0	6.40	11.72
30	0.71	710.50	269.50	220.50	3.50	4	0.00	0	6.37	11.29
31	0.71	710.50	269.50	220.50	3.50	5	0.00	0	6.40	11.67

Рис. 3.1. Вікно графічного інтерфейсу Mito для швидкої роботи з наборами даних із параметрами енергоефективності індивідуальних будинків

Перегляд даних за окремими їх атрибутами інтуїтивно зрозумілий. Натискаємо на стовпець, потім на значок фільтра та на Summary.

Встановлено, що атрибут відносна компактність індивідуальних будинків (X1 Relative Compactness) має характеристики, які подано у таблиці 3.1.

Таблиці 3.1. Характеристики атрибуту відносна компактність індивідуальних будинків (X1 Relative Compactness)

Показник	Значення
count	768.0
mean	0.76
std	0.11
min	0.62
25%	0.68
50%	0.75
75%	0.82
max	0.98
count: NaN	0
median	0.75
sum	586.66

Тут показано розподіл даних та дещо з описової статистики: середнє значення, стандартне відхилення, кількість null тощо. Атрибут відносна компактність індивідуальних будинків (X1 Relative Compactness) описується розподілом, який представлено на рис. 3.1.

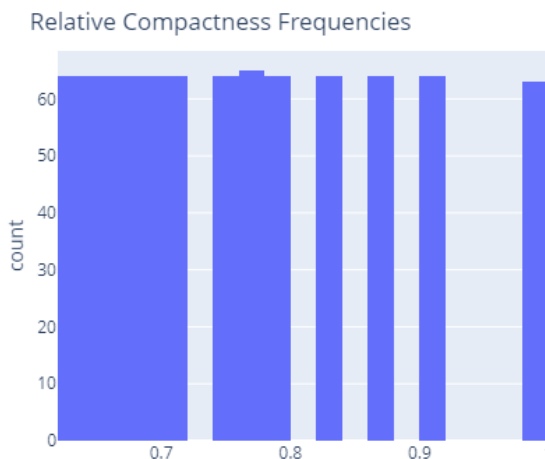


Рис. 3.1. Розподіл кількісних значень відносної компактності індивідуальних будинків

Встановлено, що атрибут площа поверхні (m^2) індивідуальних будинків (X2 Surface Area) має характеристики, які подано у таблиці 3.2.

Таблиці 3.2. Характеристики атрибуту площа поверхні (m^2) індивідуальних будинків (X2 Surface Area)

Показник	Значення
count	768.0
mean	671.71
std	88.09
min	514.5
25%	606.38
50%	673.75
75%	741.12
max	808.5
count: NaN	0
median	673.75
sum	515872.0

Атрибут площа поверхні (m^2) індивідуальних будинків (X2 Surface Area) описується розподілом, який представлено на рис. 3.2.

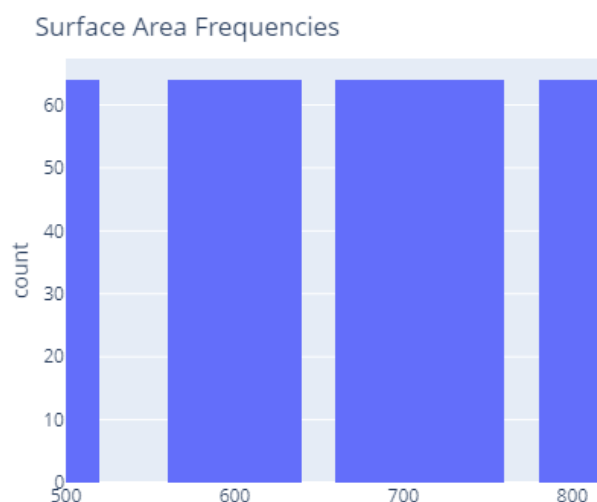


Рис. 3.2. Розподіл кількісних значень площі поверхні (м²) індивідуальних будинків (X2 Surface Area)

Встановлено, що атрибут площа стін (м²) індивідуальних будинків (X3 Wall Area) має характеристики, які подано у таблиці 3.3.

Таблиці 3.3. Характеристики атрибуту площа стін (м²) індивідуальних будинків (X3 Wall Area)

Показник	Значення
count	768.0
mean	318.5
std	43.63
min	245.0
25%	294.0
50%	318.5
75%	343.0
max	416.5
count: NaN	0
median	318.5
sum	244608.0

Атрибут площа стін (м²) індивідуальних будинків (X3 Wall Area) описується розподілом, який представлено на рис. 3.3.

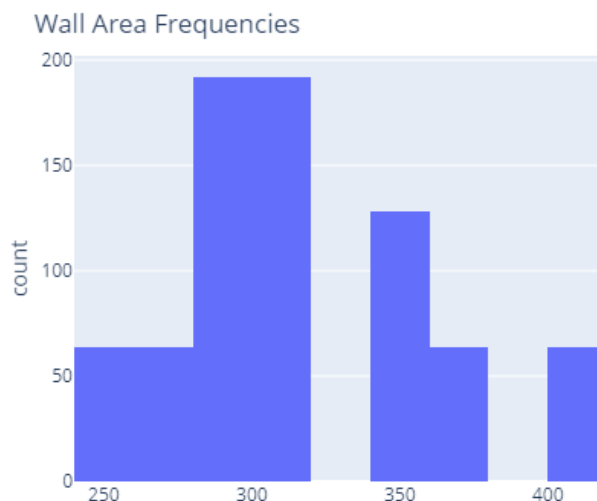


Рис. 3.3. Розподіл кількісних значень площі стін (м²) індивідуальних будинків (X3 Wall Area)

Встановлено, що атрибут площа даху (м²) індивідуальних будинків (X4 Roof Area) має характеристики, які подано у таблиці 3.4.

Таблиці 3.4. Характеристики атрибуту площа даху (м²) індивідуальних будинків (X4 Roof Area)

Показник	Значення
count	768.0
mean	176.6
std	45.17
min	110.25
25%	140.88
50%	183.75
75%	220.5
max	220.5
count: NaN	0
median	183.75
sum	135632.0

Атрибут площа даху (м²) індивідуальних будинків (X4 Roof Area) описується розподілом, який представлено на рис. 3.4.

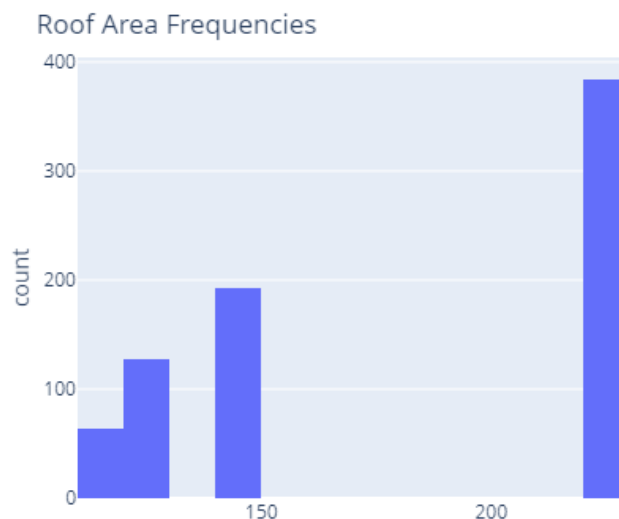


Рис. 3.4. Розподіл кількісних значень площі даху (м²) індивідуальних будинків (X4 Roof Area)

Встановлено, що атрибут загальна висота (м) індивідуальних будинків (X5 Overall Height) має характеристики, які подано у таблиці 3.5.

Таблиці 3.5. Характеристики атрибуту загальна висота (м) індивідуальних будинків (X5 Overall Height)

Показник	Значення
count	768.0
mean	5.25
std	1.75
min	3.5
25%	3.5
50%	5.25
75%	7.0
max	7.0
count: NaN	0
median	5.25
sum	768.0

Атрибут загальна висота (м) індивідуальних будинків (X5 Overall Height) описується розподілом, який представлено на рис. 3.5.

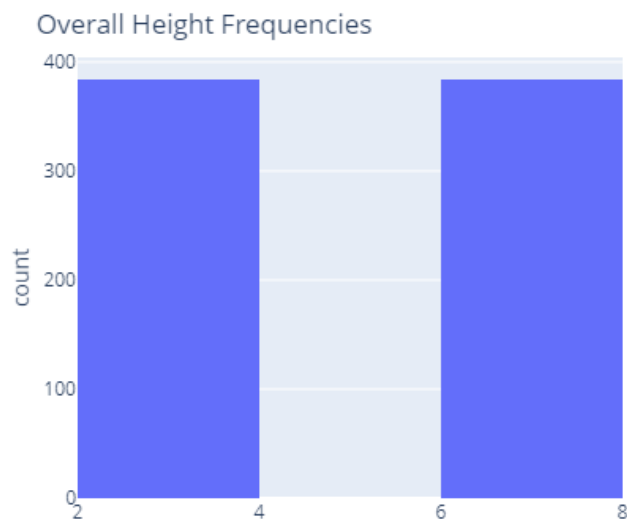


Рис. 3.5. Розподіл кількісних значень загальної висоти (м) індивідуальних будинків (X5 Overall Height)

Встановлено, що атрибут орієнтація (2: Північ, 3: Схід, 4: Південь, 5: Захід) індивідуальних будинків (X6 Orientation) має характеристики, які подано у таблиці 3.6.

Таблиці 3.6. Характеристики атрибуту орієнтація (2: Північ, 3: Схід, 4: Південь, 5: Захід) індивідуальних будинків (X6 Orientation)

Показник	Значення
count	768.0
mean	3.5
std	1.12
min	2.0
25%	2.75
50%	3.5
75%	4.25
max	5.0
count: NaN	0
median	3.5
sum	2688

Атрибут орієнтація (2: Північ, 3: Схід, 4: Південь, 5: Захід) індивідуальних будинків (X6 Orientation) описується розподілом, який представлено на рис. 3.6.

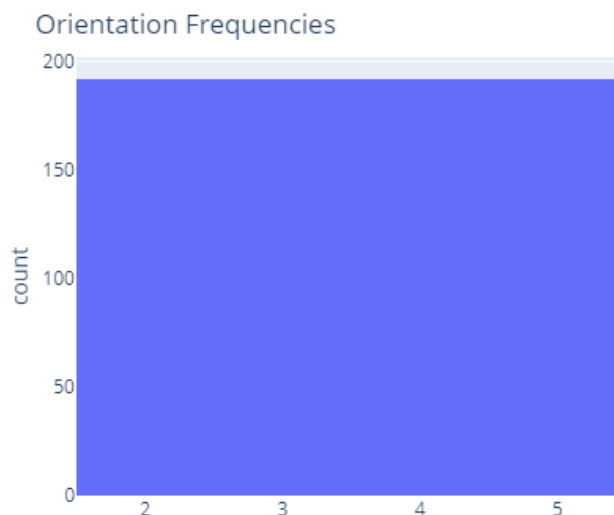


Рис. 3.6. Розподіл кількісних значень орієнтації (2: Північ, 3: Схід, 4: Південь, 5: Захід) індивідуальних будинків (X6 Orientation)

Встановлено, що атрибут площа скління (0%, 10%, 25%, 40% площі підлоги) індивідуальних будинків (X7 Glazing Area) має характеристики, які подано у таблиці 3.7.

Таблиці 3.7. Характеристики атрибуту площа скління (0%, 10%, 25%, 40% площі підлоги) індивідуальних будинків (X7 Glazing Area)

Показник	Значення
count	768.0
mean	0.23
std	0.13
min	0.0
25%	0.1
50%	0.25
75%	0.4
max	0.4
count: NaN	0
median	0.25
sum	180.0

Атрибут площа скління (0%, 10%, 25%, 40% площі підлоги) індивідуальних будинків (X7 Glazing Area) описується розподілом, який представлено на рис. 3.7.

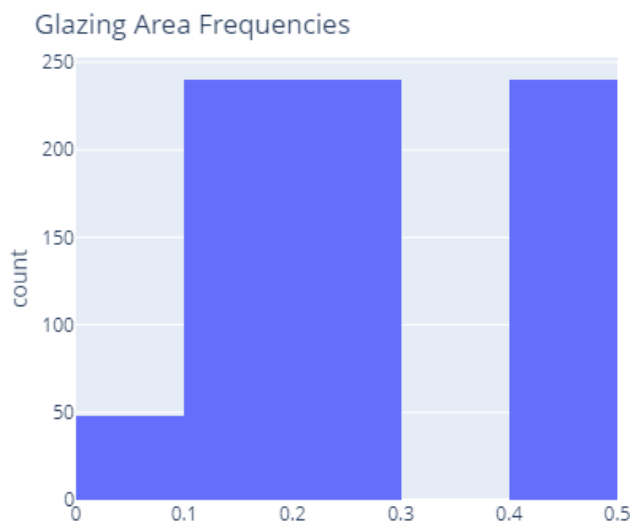


Рис. 3.7. Розподіл кількісних значень площі скління (0%, 10%, 25%, 40% площі підлоги) індивідуальних будинків (X7 Glazing Area)

Встановлено, що атрибут розподіл площі скління (відхилення) (1: Рівномірний, 2: Північ, 3: Схід, 4: Південь, 5: Захід) індивідуальних будинків (X8 Glazing Area Distribution) має характеристики, які подано у таблиці 3.8.

Таблиці 3.8. Характеристики атрибуту розподіл площі скління (відхилення) (1: Рівномірний, 2: Північ, 3: Схід, 4: Південь, 5: Захід) індивідуальних будинків (X8 Glazing Area Distribution)

Показник	Значення
count	768.0
mean	2.81
std	1.55
min	0.0
25%	1.75
50%	3.0
75%	4.0
max	5.0
count: NaN	0
median	3.0
sum	2160

Атрибут розподіл площі скління (відхилення) (1: Рівномірний, 2: Північ, 3: Схід, 4: Південь, 5: Захід) індивідуальних будинків (X8 Glazing Area Distribution) описується розподілом, який представлено на рис. 3.8.

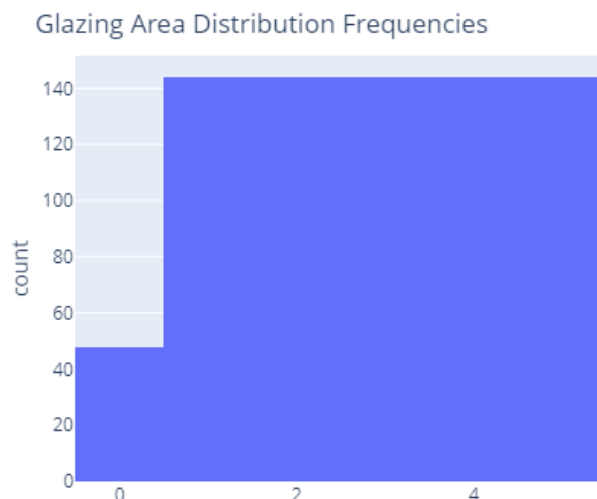


Рис. 3.8. Розподіл площі скління (відхилення) (1: Рівномірний, 2: Північ, 3: Схід, 4: Південь, 5: Захід) індивідуальних будинків (X8 Glazing Area Distribution)

Встановлено, що атрибут навантаження на опалення (кВт.год/м²) індивідуальних будинків (Y1 Heating Load) має характеристики, які подано у таблиці 3.9.

Таблиці 3.9. Характеристики атрибуту навантаження на опалення (кВт.год/м²) індивідуальних будинків (Y1 Heating Load)

Показник	Значення
count	768.0
mean	22.31
std	10.09
min	6.01
25%	12.99
50%	18.95
75%	31.67
max	43.1
count: NaN	0
median	18.95
sum	17131.93

Атрибут навантаження на опалення (кВт.год/м²) індивідуальних будинків (Y1 Heating Load) описується розподілом, який представлено на рис. 3.9.

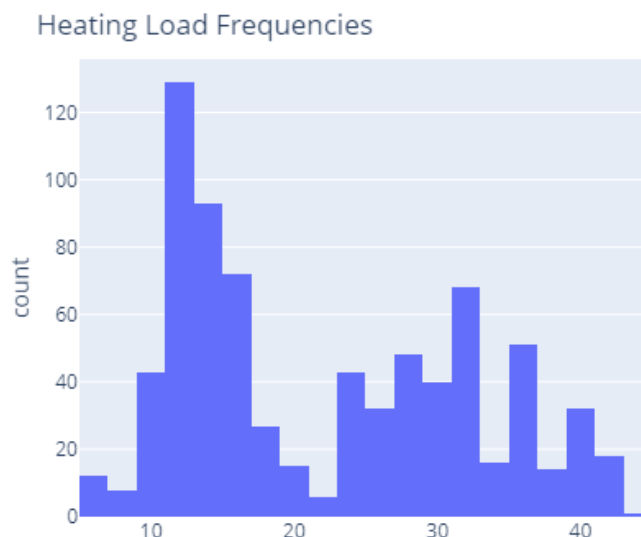


Рис. 3.9. Розподіл навантаження на опалення (кВт.год/м²) індивідуальних будинків (Y1 Heating Load)

Встановлено, що атрибут навантаження на охолодження (кВт.год/м²) індивідуальних будинків (Y2 Cooling Load) має характеристики, які подано у таблиці 3.10.

Таблиці 3.10. Характеристики атрибуту навантаження на охолодження (кВт.год/м²) індивідуальних будинків (Y2 Cooling Load)

Показник	Значення
count	768.0
mean	24.59
std	9.51
min	10.9
25%	15.62
50%	22.08
75%	33.13
max	48.03
count: NaN	0
median	22.08
sum	18883.4

Атрибут навантаження на охолодження (кВт.год/м²) індивідуальних будинків (Y2 Cooling Load) описується розподілом, який представлено на рис. 3.10.

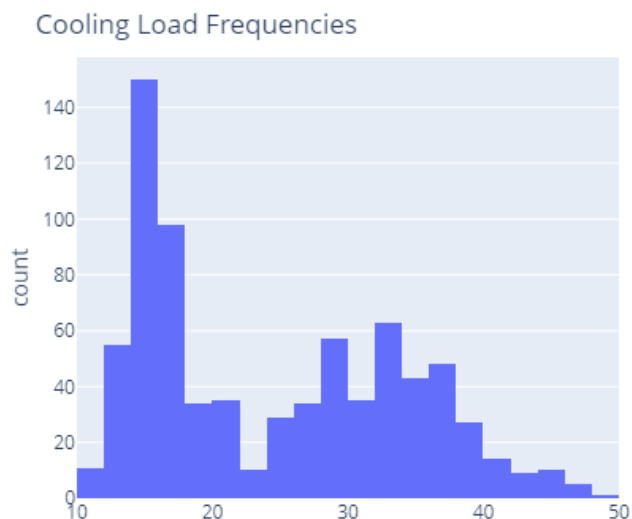


Рис. 3.10. Розподіл навантаження на охолодження (кВт.год/м²) індивідуальних будинків (Y2 Cooling Load)

Отримані результати аналізу даних лежить в основі побудови залежностей між ними та оцінення взаємозв'язків, які забезпечують формування потрібних знань для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.

3.2. Оцінення взаємозв'язків між даними для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками

На підставі проведеного аналізу даних щодо показників вимог до навантаження на системи опалення та охолодження (тобто енергоефективності) індивідуальних будинків побудовано залежності між ними та виконано оцінення взаємозв'язків, які лежать в основі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками. Зокрема, встановлено взаємозв'язок між навантаженням на охолодження (кВт.год/м² – Cooling Load) та площею поверхні (м² – X2 Surface Area), що представлено на рис. 3.11.

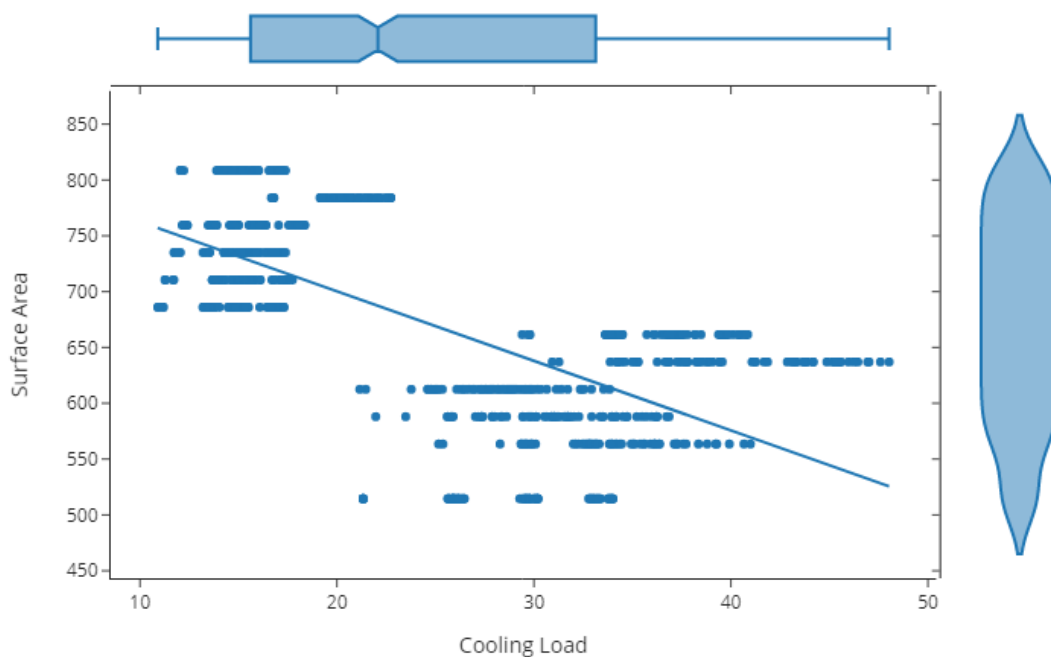


Рис. 3.11. Взаємозв'язок між навантаженням на охолодження (кВт.год/м² – Cooling Load) та площею поверхні (м² – Surface Area) індивідуальних будинків

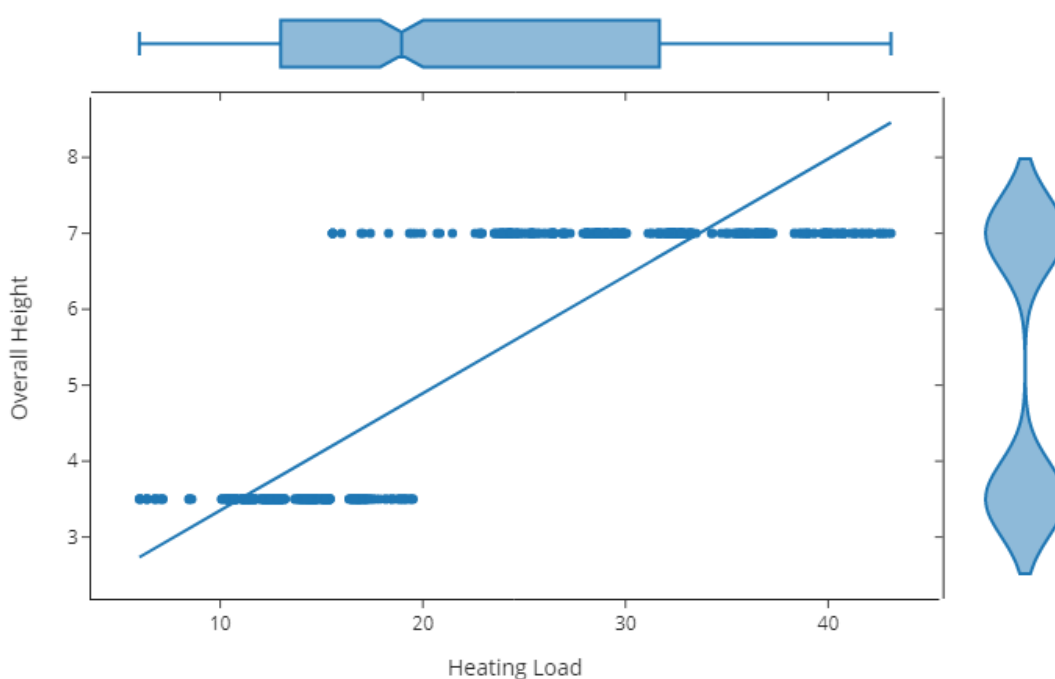


Рис. 3.12. Взаємозв'язок між навантаженням на опалення (кВт.год/м² – Heating Load) та загальною висотою (м – Overall Height) індивідуальних будинків

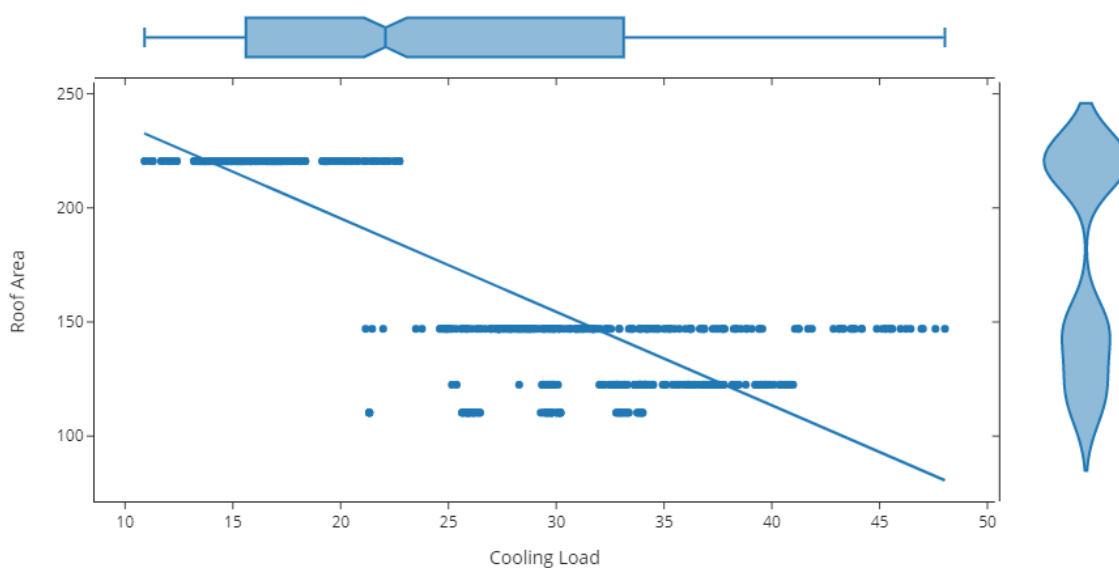


Рис. 3.13. Взаємозв'язок між навантаженням на охолодження (кВт.год/м² – Cooling Load) та площею даху (м² – X4 Roof Area) індивідуальних будинків

На підставі отриманих взаємозв'язків (рис. 3.11-3.12), які лежать в основі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками встановлено, що зв'язок між навантаженням на охолодження (кВт.год/м² – Cooling Load) та площею поверхні (м² – Surface Area) індивідуальних будинків є від'ємним і має коефіцієнт кореляції – 0,67. Водночас, зв'язок між навантаженням на опалення (кВт.год/м² – Heating Load) та загальною висотою (м – Overall Height) індивідуальних будинків є додатнім та має коефіцієнт кореляції – 0,89. Стосовно зв'язку між навантаженням на охолодження (кВт.год/м² – Cooling Load) та площею даху (м² – X4 Roof Area) індивідуальних будинків, то він також є від'ємним та має коефіцієнт кореляції – 0,86.

Матриця всіх отриманих взаємозв'язків між окремими атрибутами аналізованого набору даних показників щодо вимог до навантаження на системи опалення та охолодження (тобто енергоефективності) індивідуальних будинків представлена на рис. 3.13.

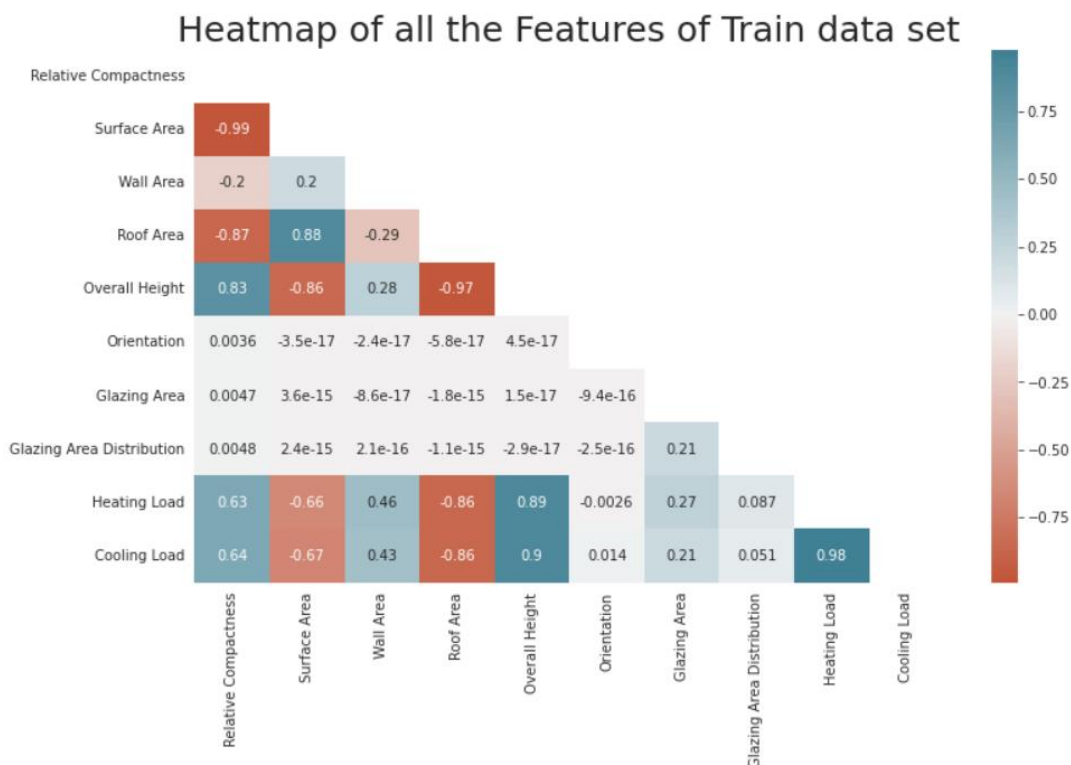


Рис. 3.13. Матриця взаємозв'язків між окремими атрибутами набору даних із показників щодо вимог до навантаження на системи опалення та охолодження (тобто енергоефективності) індивідуальних будинків

На підставі отриманої матриці взаємозв'язків між окремими атрибутами набору даних із показників щодо вимог до навантаження на системи опалення та охолодження (тобто енергоефективності) індивідуальних будинків встановлено, що більшість із них має вплив на енергоефективність. Для створення ефективною моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками слід використовувати методи машинного навчання та проведення відповідного дослідження.

3.3. Вибір інструментарію та підготовка даних до машинного навчання

На підставі аналізу набору даних із показниками вимог до навантаження на системи опалення та охолодження (тобто енергоефективності) індивідуальних будинків із використанням методів машинного навчання можливо створити модель прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками. Іншими методами виконати прогнозування попиту на електроенергію складно через множину чинників, які відображають погодні умови, фізичні характеристики будівлі та вимоги мешканців до електроенергії.

Побудова моделей прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками може бути виконана із використанням двох основних їх груп: математичні моделі та моделі, які базуються на даних. Математичні моделі, також відомі як такі, що використовують підходи до моделювання на основі опису законів фізики, а також часто вимагають багато зрозумілих вхідних даних про структуру системи та її проектне середовище. Проектне середовище містить інформацію про систему генерування та споживання електроенергії індивідуальними будинками (опалення, вентиляції та кондиціонування повітря), товщину ізоляції, теплові властивості, внутрішні навантаження на людей, дані про погодні умови регіону та багато іншого. Ці моделі потребують дослідження багатьох часто недосяжних параметрів. Через значні витрати часу на їх побудову та недостатню кількість необхідної вхідної інформації зазначені моделі є неефективними.

Машинне навчання виконуємо із використанням блокноту Jupyter Notebook. З метою обґрунтування ефективної моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками використовуємо відомі бібліотеки Python – бібліотека `scipy.stats` статистичні функції, бібліотека `CatBoost` забезпечує структуру посилення градієнта,

бібліотека Scikit-learn надає функціональність для створення та тренування різноманітних алгоритмів машинного навчання та Keras відкрита нейромережна бібліотека. Виконуємо імпорт зазначених бібліотек із потрібними модулями.

```

from scipy.stats import randint as sp_randint
from catboost import CatBoostRegressor
from sklearn.model_selection import GridSearchCV
from keras.layers import Dense
from keras.models import Sequential
from sklearn.tree import DecisionTreeRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.neural_network import MLPRegressor
from sklearn.ensemble import GradientBoostingRegressor, AdaBoostRegressor, BaggingRegressor, RandomForestRegressor
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import train_test_split
from sklearn.multioutput import MultiOutputRegressor
from sklearn.preprocessing import MinMaxScaler
from sklearn.svm import SVC
from sklearn.svm import SVR
from sklearn.metrics import accuracy_score, f1_score
from sklearn.metrics import r2_score
from sklearn.metrics import roc_auc_score

```

Після імпорту потрібних бібліотек виконуємо опис потрібних змінних та цільових функцій. Змінними є відносна компактність – Relative Compactness, площа поверхні, м² – Surface Area, площа стін, м² – Wall Area, площа даху, м² – Roof Area, загальна висота, м – Overall Height, орієнтація (2: Північ, 3: Схід, 4: Південь, 5: Захід) – Orientation, площа скління (0%, 10%, 25%, 40% площі підлоги) – Glazing Area та розподіл площі скління (відхилення) (1: Рівномірний, 2: Північ, 3: Схід, 4: Південь, 5: Захід) – Glazing Area Distribution. До цільових функцій належать навантаження на опалення, кВт.год/м² – Y1 Heating Load та навантаження на охолодження, кВт.год/м² – Y2 Cooling Load.

```

X = data[['Relative Compactness', 'Surface Area', 'Wall Area', 'Roof Area', 'Overall Height', 'Orientation', 'Glazing Area', 'Glazing Area Distribution']]
Y = data[['Heating Load', 'Cooling Load']]
Y1= data[['Heating Load']]
Y2= data[['Cooling Load']]

```

Після цього виконуємо поділ набору даних на навчальний і тестовий набір.

```
X_train, X_test, y1_train, y1_test, y2_train, y2_test = train_test_split(X, Y1, Y2, test_size=0.33, random_state = 20)

MinMax = MinMaxScaler(feature_range= (0,1))
X_train = MinMax.fit_transform(X_train)
X_test = MinMax.transform(X_test)
```

Масштабування ознак або нормалізація даних – це метод, який використовується для нормалізації діапазону незалежних змінних або характеристик даних. Отже, коли значення незалежної змінної сильно відрізняються, ми використовуємо масштабування ознак, щоб усі значення залишалися в порівнянному діапазоні.

3.4. Наближене оцінення моделей прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками

Для наближеного оцінення моделей прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками, з-поміж яких вибираються базові моделі для подальших досліджень, насамперед створено DataFrame для зберігання результатів обчислень, отриманих за допомогою різних розглядуваних моделей.

```
Acc = pd.DataFrame(index=None, columns=['model', 'train_Heating', 'test_Heating', 'train_Cooling', 'test_Cooling'])
```

На наступному етапі вибрано декілька регресійних моделей, щоб перевірити їх продуктивність.


```

regressors = [['SVR',SVR()],

              ['DecisionTreeRegressor',DecisionTreeRegressor()],
              ['KNeighborsRegressor', KNeighborsRegressor()],
              ['RandomForestRegressor', RandomForestRegressor()],
              ['MLPRegressor',MLPRegressor()],
              ['AdaBoostRegressor',AdaBoostRegressor()],
              ['GradientBoostingRegressor',GradientBoostingRegressor()]]

```

Виконуємо оцінення моделей прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.

```

for mod in regressors:
    name = mod[0]
    model = mod[1]

    model.fit(X_train,y1_train)
    actr1 = r2_score(y1_train, model.predict(X_train))
    acte1 = r2_score(y1_test, model.predict(X_test))

    model.fit(X_train,y2_train)
    actr2 = r2_score(y2_train, model.predict(X_train))
    acte2 = r2_score(y2_test, model.predict(X_test))

    Acc = Acc.append(pd.Series({'model':name, 'train_Heating':actr1,'test_Heating':acte1,'train_Cooling':actr2,'test_Cooling':acte2}),ignore_index=True )
Acc.sort_values(by='test_Cooling')

```

У результаті моделювання отримано можливість виміряти точність вибраних моделей машинного навчання для прогнозування навантаження на опалення та охолодження індивідуальних будинків.

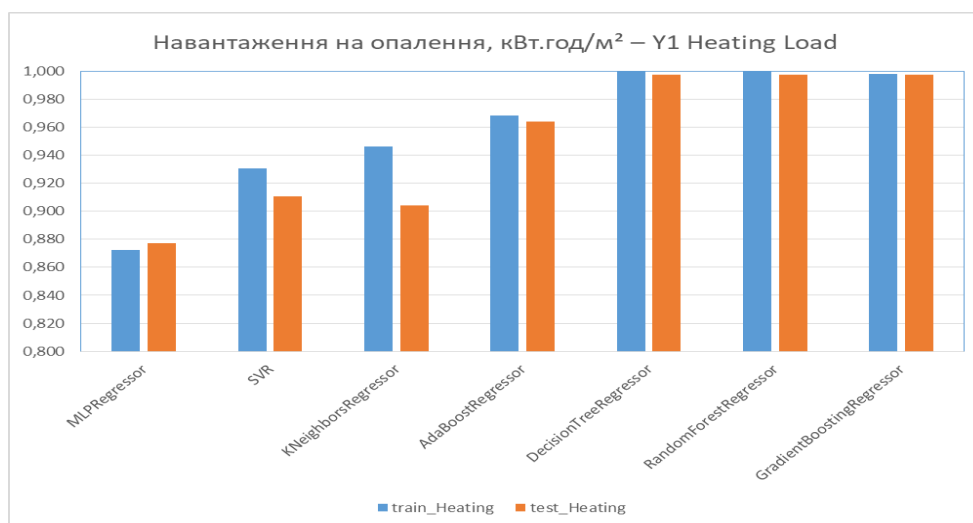


Рис. 3.14. Результати співставлення навчальних та тестових наборів стосовно прогнозування навантаження на опалення

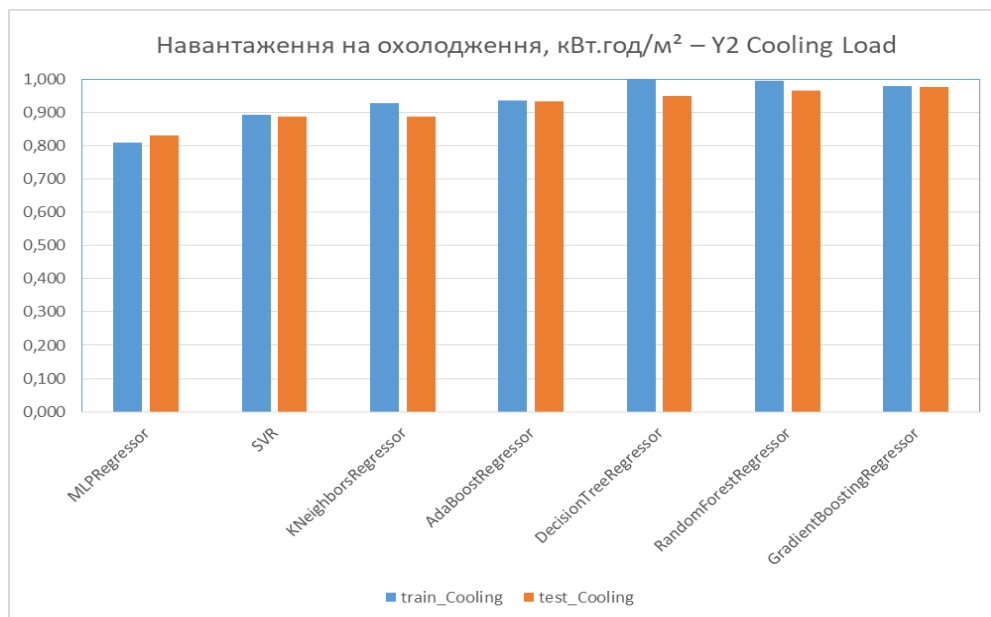


Рис. 3.15. Результати співставлення навчальних та тестових наборів стосовно прогнозування навантаження на охолодження

Отримані результати досліджень свідчать про те, що найбільш точними для прогнозування навантаження на опалення та охолодження індивідуальних будинків є чотири моделі, які базуються на алгоритмах DecisionTreeRegressor, RandomForestRegressor, MLPRegressor та GradientBoostingRegressor, які потребують подальших досліджень. При цьому найбільш точною є модель, яка базується на алгоритмі GradientBoostingRegressor.

3.5. Результати обґрунтування раціональної моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками

Насамперед виконаємо налаштування параметрів моделей. Вибрані для подальших досліджень алгоритми машинного навчання широко використовуються, оскільки вони забезпечують кращу точність порівняно із простими. Продуктивність цих алгоритмів залежить від гіперпараметрів.

Оптимальний набір параметрів може допомогти досягти більш високої точності прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками. Знаходження гіперпараметрів вручну – справа досить трудомістка та дорога. Тому автоматизація налаштування гіперпараметрів є важливою. RandomSearch, GridSearchCV і байєсовська оптимізація зазвичай використовуються для оптимізації гіперпараметрів.

У нашій роботі визначаємо раціональні параметри для моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками за допомогою алгоритму «GridSearchCV».

Алгоритм «Decision Tree Regressor» став одним із найбільш використовуваних алгоритмів машинного навчання як у навчання, так і у бізнес-середовищі. Decision Tree Regressor можна використовувати як у задачі класифікації, так і в задачі регресії. Модель із використанням зазначеного алгоритму базується на правилах прийняття рішень, отриманих із навчальних даних. У задачі регресії модель використовує значення замість класу, а середня квадратична помилка використовується для оцінення точності рішення. Модель дерева рішень «Decision Tree Regressor» погано підходить для узагальнення та чутлива до змін навчальних даних. Невелика зміна в навчальному наборі даних може вплинути на точність прогнозування моделі.

До основних параметрів моделі на базі алгоритму «Decision Tree Regressor» належать:

- **max_features**: кількість випадково вибраних функцій, з яких потрібно вибрати найкращу функцію для поділу на заданий вузол дерева. Це може бути ціле число або один із двох наведених нижче методів (auto: квадратний корінь із загальної кількості предикторів. max: кількість предикторів.);
- **max_leaf_nodes**: максимальна кількість листових вузлів, яку може мати дерево в лісі, ціле число від 1 до $1e9$ включно;
- **max_depth**: максимальна глибина для вирощування кожного дерева, ціле число від 1 до 100 включно;

➤ **min_samples_leaf**: мінімальна кількість зразків, яку має мати кожна гілка після поділу вузла, ціле число від 1 до $1e6$ включно. Поділ, який викликає менше зразків, що залишилися, відкидається.

Виконаємо налаштування параметрів моделі за допомогою навчального набору даних Y2 Cooling Load (або `y2_train`).

```
DTR = DecisionTreeRegressor()
param_grid = {"criterion": ["mse", "mae"], "min_samples_split": [14, 15, 16, 17],
              "max_depth": [5, 6, 7], "min_samples_leaf": [4, 5, 6], "max_leaf_nodes": [29, 30, 31, 32],}

grid_cv_DTR = GridSearchCV(DTR, param_grid, cv=5)

grid_cv_DTR.fit(X_train, y2_train)
print("R-Squared: {}".format(grid_cv_DTR.best_score_))
print("Best Hyperparameters: \n{}".format(grid_cv_DTR.best_params_))
```

У результаті виконання моделювання після обґрунтування параметрів моделей встановлено, що точність вибраних моделей машинного навчання для прогнозування навантаження на опалення та охолодження індивідуальних будинків дещо покращилася (рис. 3.16-3.17).

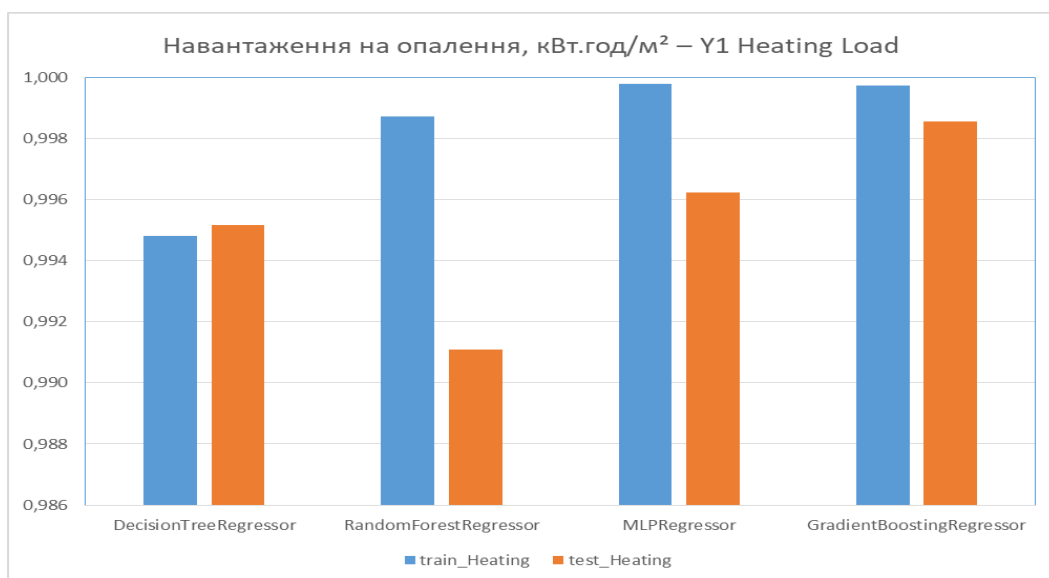


Рис. 3.16. Результати співставлення навчальних та тестових наборів стосовно прогнозування навантаження на опалення після обґрунтування раціональних параметрів моделей

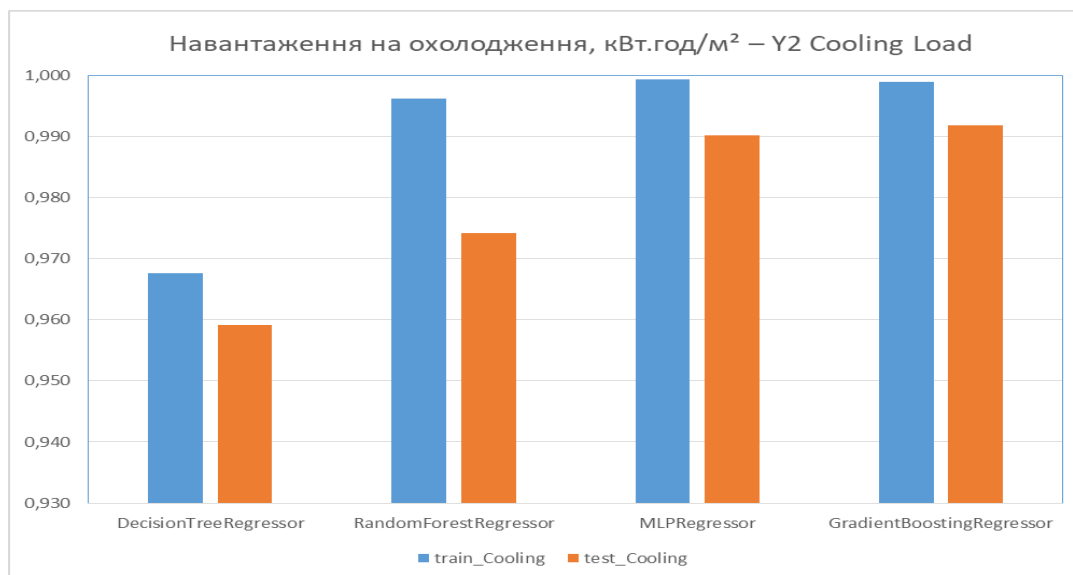


Рис. 3.17. Результати співставлення навчальних та тестових наборів стосовно прогнозування навантаження на охолодження

Встановлено, що найвищу точність забезпечує модель прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками, яка базується на алгоритмі GradientBoostingRegressor.

У подальшому виконаємо навчання моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками, яка базується на алгоритмі GradientBoostingRegressor.

```

model = CatBoostRegressor(border_count= 5, depth= 8, iterations= 10000, learning_rate= 0.02,
random_state= 45)

model.fit(X_train,y1_train)
actr1 = r2_score(y1_train, model.predict(X_train))
acte1 = r2_score(y1_test, model.predict(X_test))
y1_pred = model.predict(X_test)

model.fit(X_train,y2_train)
actr2 = r2_score(y2_train, model.predict(X_train))
acte2 = r2_score(y2_test, model.predict(X_test))
y2_pred = model.predict(X_test)

```

Запропоновану модель на основі алгоритму GradientBoostingRegressor можна використовувати для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками. Слід провести перевірку точності виконаного прогнозу обсягів генерування та споживання електроенергії

індивідуальними будинками із порівнянням отриманих значень прогнозу та відомих історичних даних (рис. 3.18).

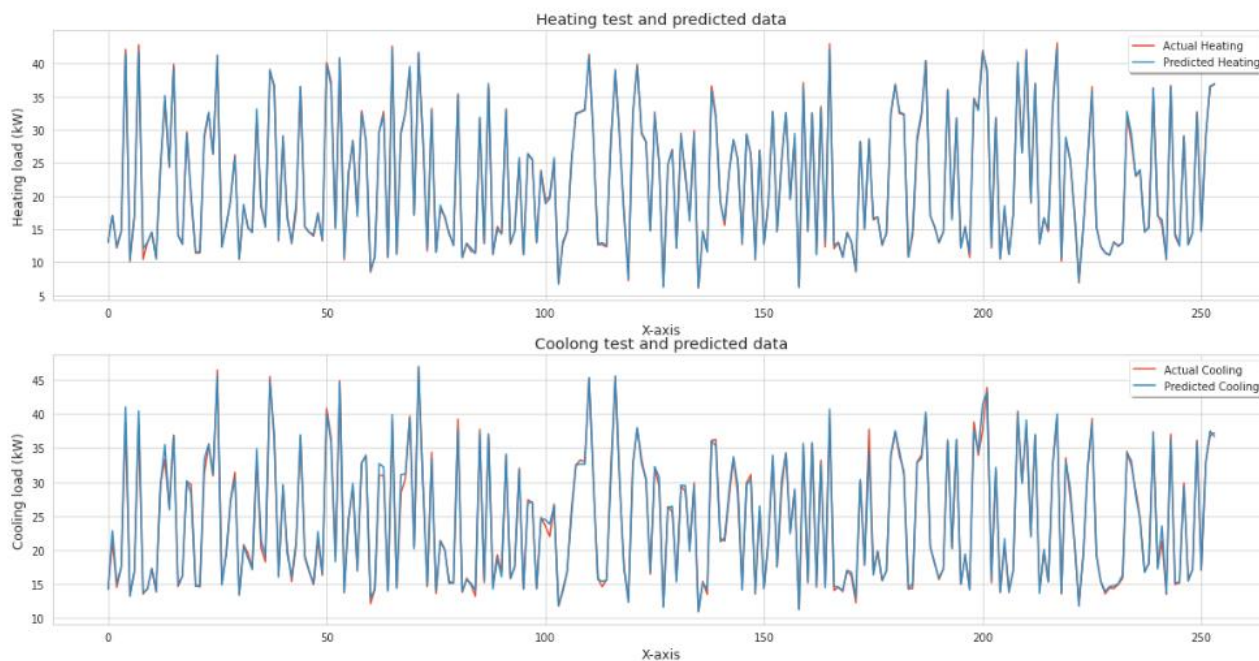


Рис. 3.18. Результати перевірки точності виконаного прогнозу обсягів генерування та споживання електроенергії індивідуальними будинками із порівнянням отриманих значень прогнозу та відомих історичних даних

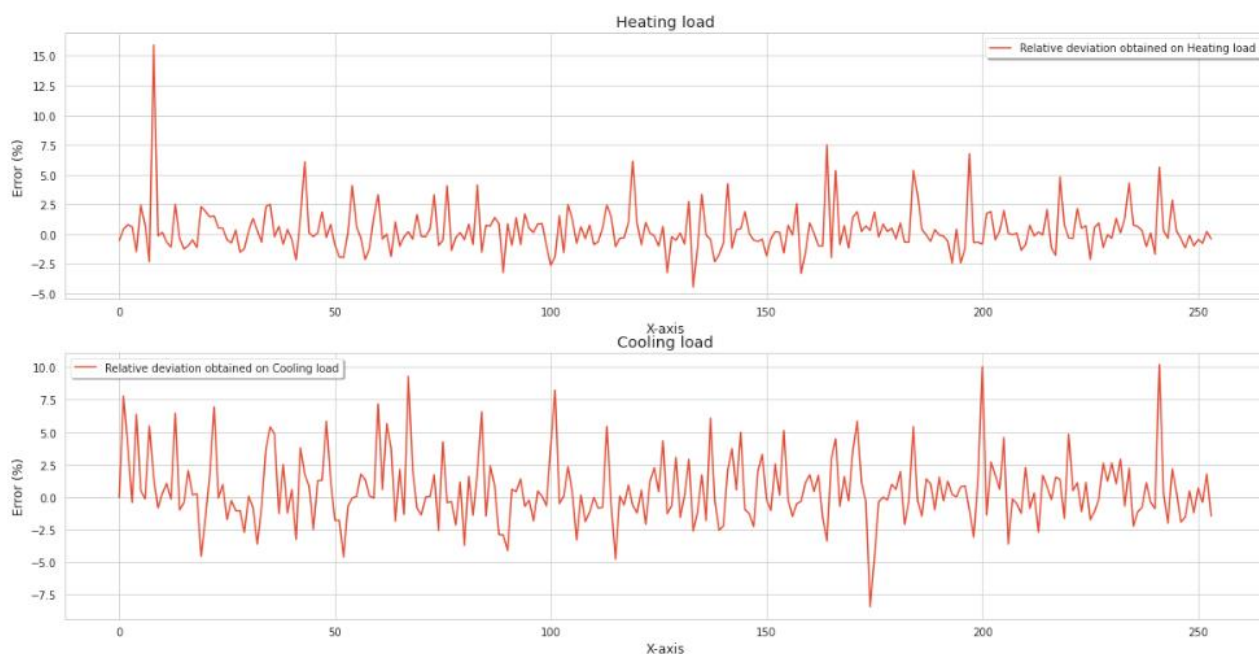


Рис. 3.19. Результати прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками

Точність виконаного прогнозу перевіряємо на підставі розрахунку коефіцієнт детермінації (табл. 3.1).

Таблиця 3.1. Результати перевірки точності прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками із використання алгоритму GradientBoostingRegressor

Показник	Варіант досліджу	train dataset	test dataset
Коефіцієнт детермінації R-Squared	1	0.9999	0.9988
	2	0.9999	0.9943

При цьому після налаштування різних параметрів в моделі отримали дуже хороший результат прогнозу обсягів генерування та споживання електроенергії індивідуальними будинками (>99,5% як для навантажень опалення, так і для охолодження, порівняно з набором експериментальних даних).

РОЗДІЛ 4.

ОХОРОНА ПРАЦІ ТА БЕЗПЕКА У НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1. Аналіз небезпечних та шкідливих виробничих чинників під час роботи з комп'ютерною технікою

Працівники, задіяні на роботах, пов'язаних з періодичною або постійною роботою за комп'ютером, піддаються впливу факторів виробничої безпеки, основними з яких є:

1. Фізичні

- Підвищений рівень напруги в електричному ланцюзі, замикання якої може пройти через тіло працюючого.
- Підвищений рівень рентгенівського випромінювання.
- Підвищений рівень ультрафіолетового випромінювання.
- Підвищений рівень інфрачервоного випромінювання.
- Можливість ураження статичною електрикою.
- Запиленість повітря робочого приміщення.
- Підвищений вміст важких (+) аероіонів.
- Нерівномірний розподіл яскравості в полі зору.
- Підвищений рівень пульсації світлового потоку.

2. Хімічні.

- Підвищений вміст у повітрі вуглекислого газу, озону, аміаку, фенолу, формальдегіду та інших речовин.

3. Психофізіологічні.

- Напруга зору.
- Напруга пам'яті.
- Напруга уваги.
- Тривале статичне напруження.
- Відносно великий обсяг інформації, що обробляється в одиницю часу.

- Монотонність праці в окремих випадках.
- Нераціональна організація робочого місця.

До основних шкідливих факторів при роботі з комп'ютером відносять: тривале сидяче положення, електромагнітне випромінювання, навантаження на зір, перевантаження кистьових суглобів, можливість захворювань органів дихання, алергії, порушення нормального перебігу вагітності та ін ^ Тривале сидяче положення приводить до напруги м'язів шиї, голови, рук і плечей, остеохондрозу, у дітей - ще й до сколіозу.

Тривале сидяче положення ще приводить до застою крові в тазових органах і, як наслідок, до простатиту й геморою. Малорухливий спосіб життя призводить до ожиріння. Остеохондроз виникає при порушенні міжхребцевих дисків, яке призводить до випинання в яку або сторону (грижі міжхребцевого диска). Грижа може зашкодити спинний мозок і нервові відростки. Наслідки можуть бути найрізноманітнішими, від болів в спині і кінцівках, до паралічу кінцівок і смерті. Одна з поширених причин остеохондрозу - дистрофія м'язів спини.

Людина, що проводить в основному сидячий спосіб життя, цілком може захворіти остеохондрозом. Ознаки початку захворювання: дискомфорт у спині та больові відчуття, головні болі, порушення роботи внутрішніх органів. До факторів ризику захворювання гемороєм відносять: сидячий спосіб життя, ожиріння, надмірне вживання копчених, гострих, солоних і пряних продуктів, запальні захворювання малого таза та ін Ожиріння виникає через нераціональне харчування, малорухомого і в тому числі сидячого способу життя, неадекватної реакції на стресові ситуації, надмірно довгий сон, застосування гормональних препаратів, перевантаження організму харчовими жирами тощо.

Ожиріння призводить до збільшення навантаження на серце, зміни конфігурації та положення серця в грудній порожнині, підвищення вмісту холестерину в крові, в результаті він відкладається на стінках судин (атеросклероз). Підвищений скупчення жиру всередині грудної порожнини

впливає на роботу органів дихання, що призводить до появи задишки та гіпоксії органів і тканин.

Навантаження на зір. Людське око реагує на найдрібнішу вібрацію тексту і на мерехтіння екрану. М'язи ока, керуючі кришталиком, перебувають у постійній нарузі, що обов'язково призводить до втрати гостроти зору. Немаловажне значення для профілактики зорових дисфункцій надають: правильний чи рекомендований підбір кольору, шрифтів, компоновки вікон у використовуваних додатках, орієнтація дисплея монітора. Тривала робота за комп'ютером – це величезне навантаження на очі, оскільки зображення на моніторі складається не з безперервних ліній, як на папері, а з окремих точок, які світяться і мерехтять. У користувача неминуче погіршується зір, очі починають слізотитися, з'являється головний біль, втома, зображення двоїться і спотворюється.

4.2. Моделювання процесу виникнення травм та аварій

Метод логічного моделювання процесів формування, виникнення небезпечних ситуацій та їх наслідків доцільно застосовувати для аналізу існуючих або потенційних небезпек, що виявленні під час обстеження робочих місць, окремих видів комп'ютерної техніки. Але, як показали дослідження, будь-яка аварія може бути наслідком однієї чи багатьох потенційних небезпечних ситуацій або їх поєднання. Тому метод логічного моделювання не може бути застосований для моделювання складних аварій і катастроф. Обчислення рівняння безпеки можна спрямувати на удосконалення конструкції комп'ютерної техніки, на зниження їх безпеки, а також вживати термінових заходів для першочергового усунення небезпек з більш високим рівнем.

Аналіз умов, обставин та причин різних аварій, виробничих травм та деяких катастроф показав, що процеси формування та виникнення цих явищ

можна заздалегідь моделювати, застосовуючи метод побудови «дерева відмов» та помилок оператора комп'ютерної техніки. Так, побудовані операторні або логіко-імітаційні моделі травм при роботі із комп'ютерною технікою.

Аналіз моделей процесів формування й виникнення аварій, травм показав, що вони повністю імітують усі процеси та явища, що беруть участь у їх зародженні й виникненні. У зв'язку з цим, моделі, що отримали назву «дерево відмов комп'ютерної техніки і помилок оператора» можна назвати імітаційними. А оскільки виникнення кожної наступної події знаходять шляхом логічного аналізу попередніх, то для кращого розуміння суті таких моделей, їх можна назвати логічно-імітаційними.

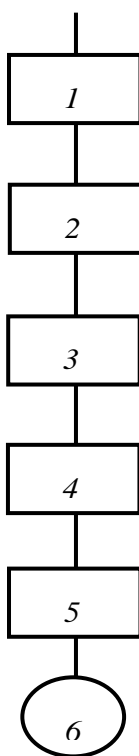


Рис. 4.1. Схема побудови логіко-імітаційної моделі:
1 – головна подія; 2-5 – проміжні події; 6 – базова подія.

Основні принципи побудови моделі такі. Виявляється комп'ютеризований процес, на якому вже були раніше або можуть статися аварії, виробничі травми чи катастрофи. За своєю формою модель нагадує крону дерева, тому вона і одержала назву «дерево відмов і помилок». Кінцеві події називають базовими.

Для побудови логіко-імітаційних моделей застосовують різні символи, що характеризують ті чи інші події. Як правило, побудова моделі починається з головної події, а наступні розміщують зверху вниз, аж до базових подій (рис. 4.1).

Кожен блок рисунку, позначений відповідним номером, означає подію або окремий етап побудови моделі:

- 1 – відмова (аварія, травма) системи – головна подія;
- 2 – послідовність подій, що приводять до відмови системи;
- 3 – послідовність подій зображується за допомогою логічних операторів;
- 4 – усі вхідні і вихідні події, що входять до моделі, зображуються у вигляді прямокутників з відповідними написами всередині;
- 5 – послідовний підхід до базових подій, частоти виникнення яких відомі;
- 6 – базові події зображують у вигляді кружечків з написами всередині, вони є межею аналізу побудованої моделі.

4.3. Розробка заходів щодо безпеки у надзвичайних ситуаціях

Забезпечення безпеки населення у надзвичайних ситуаціях під час разі загрози та виникнення надзвичайних ситуацій є одним з найважливіших завдань, яке покладається на службу з охорони праці окремої ІТ-компанії чи організації.

Захист населення базується на державній системі заходів, що забезпечують виконання організаційних, інженерно - технічних, санітарно - гігієнічних та інших заходів в сфері запобігання та ліквідації наслідків надзвичайних ситуацій.

До надзвичайних ситуацій природного характеру, які можуть виникнути на території господарства належить: пожежа, ураган, смерч, землетрус, великі опади дощів.

З метою захисту населення, зменшення втрат та шкоди економіці в разі виникнення надзвичайних ситуацій має проводитись спеціальний комплекс заходів, а саме:

- оповіщення та інформування населення про надзвичайну ситуацію, яка може виникнути;
- спостереження і контроль за довкіллям, продуктами харчування і водою;
- створення захисних споруд та укриття в них усього населення відповідно до приналежності (працююча зміна, населення, яке проживає в небезпечній зоні);
- проведення медичного захисту для зменшення ступеня ураження людей, своєчасне надання допомоги та лікування.

РОЗДІЛ 5.

ВИЗНАЧЕННЯ ЕФЕКТИВНОСТІ ВІД ВИКОРИСТАННЯ МОДЕЛІ ПРОГНОЗУВАННЯ ОБСЯГІВ ГЕНЕРУВАННЯ ТА СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ ІНДИВІДУАЛЬНИМИ БУДИНКАМИ

Науковці визначають кількісні та якісні показники використання інформаційних технологій у діяльності окремих підприємств. Так, такі вчені як Плєскач В.Л., Рогущина Ю.В., Кустова Н.П. у своїй роботі [8] зазначають, що під час аналізу інформаційних технологій доцільно використовувати критерій зв'язку між витратами та задоволенням інформаційних потреб користувачів:

$$K = W - Z, \quad (5.1)$$

де W – вартісне оцінення рівня задоволення інформаційними потребами користувачів; Z – витрати коштів на розроблення, а також функціонування інформаційної технології обробки прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками:

$$Z = Z_r \cdot E + Z_e + Z_m, \quad (5.2)$$

де Z_r – витрати на розроблення моделі, налагодження та впровадження прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками; E – коефіцієнт економічної ефективності капіталовкладень; Z_e – експлуатаційні затрати, які пов'язані із прогнозуванням обсягів генерування та споживання електроенергії індивідуальними будинками; Z_m – витрати на адаптацію запропонованої моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.

Важливим вважають показники функціональної повноти (F), які показують відношення обсягу автоматизованого оброблення інформації (Q_a)

до обсягу оброблення інформації під час прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками Q_i :

$$F = Q_a / Q_i, \quad (5.3)$$

де F – показник функціональної повноти моделі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками; Q_a – обсяг автоматизованого оброблення інформації; Q_i – сумарний обсяг оброблення інформації під час прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками.

Розрахунок проведемо на прикладі прогнозування обсягів генерування та споживання електроенергії на території окремої громади, яка нараховує 325 індивідуальних будинків. При цьому вартісне оцінення рівня задоволення інформаційними потребами користувачів становить $W=83500$ грн.

Витрати коштів Z на розроблення, а також функціонування інформаційної технології обробки прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками становлять:

$$Z = 15600 \cdot 1,1 + 7800 + 8200 = 33160 \text{ грн.}$$

Підставивши відповідні значення у формулу (1) отримаємо:

$$K = 83500 - 33160 = 50340 \text{ грн.}$$

Показник функціональної повноти (F) становить:

$$F = 89 / 100 = 0,89.$$

Отже, виконані розрахунки свідчать про те, що функціонування інформаційної технології обробки прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками для громади із 325

індивідуальних будинків окупиться за 0,6 року і прибуток становитиме 50340 грн. Показник функціональної повноти інформаційної технології обробки прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками становить 0,89.

ВИСНОВКИ І ПРОПОЗИЦІЇ

У період війни на території України, де частково пошкоджено енергетичні системи, досить актуальними є задачі планування енергоспоживання як окремими підприємствами, так і індивідуальними будинками на території громад. Метою нашої роботи є підвищення ефективності прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками завдяки обґрунтуванню раціональної моделі, що базується на алгоритмі машинного навчання.

Нами виконано аналіз стану споживання електроенергії будівлями та обґрунтування доцільності її прогнозування. Встановлено, що орієнтовно 40% усіх енергетичних ресурсів на території України споживаються будівлями. При цьому обсяги енергозбереження у окремих будівлях можуть становити орієнтовно 50...60%. Рівень споживання електроенергії індивідуальними будинками залежить від великої множини чинників. Значні витрати на електроенергію у будівлях пов'язані з поганим повітрообміном та відсутністю гарячого водопостачання, або ж не враховуються витрати енергії на гаряче водопостачання.

Проведений аналіз відомих досліджень, а також публікацій, що стосуються прогнозування обсягів генерування та споживання електроенергії, свідчить про активність дослідників у зазначеному напрямі. Для цього використовуються різні підходи та методи, у основу яких покладено аналіз ретроспективних динамік генерування та споживання електроенергії, а також врахування різних чинників впливу на їх обсягів.

На підставі вище сказано встановлено, що слід провести дослідження методів машинного навчання, які забезпечують прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками на підставі історичних даних. Для виконання наших досліджень здійснено вибір методів машинного навчання для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками. Зокрема, алгоритм

дерева рішень (LDTR) є потужним і універсальним, оскільки він дуже ефективний для фіксації взаємодії між функціями та його легко візуалізувати, що робить їх ефективними для класифікації та прогнозування даних для вирішення різних прикладних задач. Нами представлена блок-схема запропонованої моделі на основі алгоритму LDTR.

Випадковий ліс (Random Forest Regressor) являє ансамбль дерев рішень. Це означає, що багато дерев, побудованих певним «випадковим» чином, утворюють випадковий ліс. Кожне дерево створюється з різної вибірки рядків, і в кожному вузлі для розділення вибирається різна вибірка функцій. Кожне з дерев робить свій індивідуальний прогноз. Потім ці прогнози усереднюються для отримання єдиного результату.

Багатошаровий перцептрон MLP відрізняється від інших методів тим, що існує багато параметрів, які можна налаштувати та згрупувати за двома критеріями. По-перше, щоб виконати різні дослідження, а по-друге, щоб отримати кращу продуктивність та результат прогнозування. Функція активації вихідного нейрона залежить від того, який вихід ви хочете отримати. Коли використовується функція активації у вихідному нейроні, здійснюється обмеження діапазону виведення.

Підвищення градієнта (Gradient Boosting Regressor) є одним із варіантів методів ансамблю, коли ви створюєте кілька слабких моделей і комбінуйте їх, щоб отримати кращу продуктивність у цілому. Для цього слід створити дерева регресії досліджуваного об'єкта з посиленням градієнта, який має нелінійний зв'язок між x та y . У результаті отримується комбіноване прогнозування F_m досліджуваного показника, яке стає все ближчим до поставленої мети, оскільки додається більше дерев до комбінованої моделі. Означений працює градієнтного посилення для моделей прогнозування складних об'єктів виконується шляхом поєднання кількох менших моделей.

Для виконання прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками обрано загальнодоступний набір даних. Представлений набір даних містить 8 вхідних змінних ($X_1 \dots X_8$) та 2

вихідних змінних (Y_1 та Y_2). Нас цікавить цей набір даних, щоб вивчити три цікаві напрямки щодо отримання знань із даних, а також їх візуалізувати та встановити як вони впливають на енергетичне навантаження індивідуальних будинків.

Для попереднього аналізу даних використовували графічний інтерфейс для швидкої роботи з наборами даних Mito, який за окремими діями створює код на Python. Представлений набір даних із параметрами енергоефективності індивідуальних будинків опрацьовуємо із використанням окремого блокноту JupyterLab notebook. Насамперед імпортуємо бібліотеку рядком `import mitosheet`, і запускаємо її командою `mitosheet.sheet()`.

Нами оцінено характеристики вхідних атрибутів відносна компактність, площа поверхні, площа стін, площа даху, загальна висота, орієнтація, площа скління, розподіл площі скління, а також вихідних атрибутів навантаження на опалення та охолодження індивідуальних будинків, що забезпечило побудову їх розподілів.

На підставі отриманих взаємозв'язків, які лежать в основі прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками встановлено, що зв'язок між навантаженням на охолодження та площею поверхні індивідуальних будинків є від'ємним і має коефіцієнт кореляції – 0,67. Водночас, зв'язок між навантаженням на опалення та загальною висотою індивідуальних будинків є додатнім та має коефіцієнт кореляції – 0,89. Стосовно зв'язку між навантаженням на охолодження та площею даху індивідуальних будинків, то він також є від'ємним та має коефіцієнт кореляції – 0,86.

На підставі отриманої матриці взаємозв'язків між окремими атрибутами набору даних із показників щодо вимог до навантаження на системи опалення та охолодження (тобто енергоефективності) індивідуальних будинків встановлено, що більшість із них має вплив на енергоефективність.

Машинне навчання виконуємо із використанням блокноту Jupyter Notebook. З метою обґрунтування ефективної моделі прогнозування обсягів

генерування та споживання електроенергії індивідуальними будинками використовуємо відомі бібліотеки Python – бібліотека `scipy.stats` статистичні функції, бібліотека `CatBoost` забезпечує структуру посилення градієнта, бібліотека `Scikit-learn` надає функціональність для створення та тренування різноманітних алгоритмів машинного навчання та `Keras` відкрита нейромережна бібліотека.

Результати досліджень свідчать про те, що найбільш точними для прогнозування навантаження на опалення та охолодження індивідуальних будинків є чотири моделі, які базуються на алгоритмах `DecisionTreeRegressor`, `RandomForestRegressor`, `MLPRegressor` та `GradientBoostingRegressor`, які потребують подальших досліджень. При цьому найбільш точною є модель, яка базується на алгоритмі `GradientBoostingRegressor`.

У результаті виконання моделювання після обґрунтування параметрів моделей встановлено, що точність вибраних моделей машинного навчання для прогнозування навантаження на опалення та охолодження індивідуальних будинків дещо покращилася. Запропоновану модель на основі алгоритму `GradientBoostingRegressor` можна використовувати для прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками. Слід провести перевірку точності виконаного прогнозу обсягів генерування та споживання електроенергії індивідуальними будинками із порівнянням отриманих значень прогнозу та відомих історичних даних.

Точність виконаного прогнозу перевіряємо на підставі розрахунку коефіцієнт детермінації. При цьому після налаштування різних параметрів в моделі отримали дуже хороший результат прогнозу обсягів генерування та споживання електроенергії індивідуальними будинками (>99,5% як для навантажень опалення, так і для охолодження, порівняно з набором експериментальних даних).

Запропоновано заходи із охорони праці та безпеки у надзвичайних ситуаціях, дотримання яких дасть можливість створити безпечні умови праці під час роботи з комп'ютерною технікою.

Виконані розрахунки економічної ефективності свідчать про те, що функціонування інформаційної технології обробки прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками для громади із 325 індивідуальних будинків окупиться за 0,6 року і прибуток становитиме 50340 грн. Показник функціональної повноти інформаційної технології обробки прогнозування обсягів генерування та споживання електроенергії індивідуальними будинками становить 0,89.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Базовий рівень споживання енергії громадськими будівлями. Роль базового рівня при проведенні енергетичної оцінки будівель та застосуванні механізму ЕСКО [електронний ресурс]. – режим доступу: <http://slavrada.gov.ua/uploads/File/invest/posibnyki/15.pdf>.
2. Жидецький В.Ц., Джигирей В.С., Мельников О.В. Основи охорони праці. Підручник. Вид. 5-е, доповнене. Львів: Афіша, 2012. 350с.
3. Закон України «Про внесення змін до Бюджетного кодексу України щодо запровадження нових інвестиційних можливостей, гарантування прав та законних інтересів суб'єктів підприємницької діяльності для проведення масштабної енергомодернізації» [Електронний ресурс]. Верховна Рада України. Режим доступу : <http://zakon5.rada.gov.ua/laws/show/328-19>.
4. Закон України «Про запровадження нових інвестиційних можливостей, гарантування прав та законних інтересів суб'єктів підприємницької діяльності для проведення масштабної енергомодернізації» [Електронний ресурс]. Верховна Рада України. Режим доступу : <http://zakon3.rada.gov.ua/laws/show/327-19>.
5. Косухіна О.С., Тонконог С.Є. Алгоритмічне забезпечення методу прогнозування обсягів споживання електроенергії. Математичне моделювання. 1 (36), 2017. С. 11-16.
6. Лехман С.Д., Рублев В.І., Рябцев Б.І. Запобігання аварійності і травматизму у сільському господарстві. К.: Урожай, 1993. 267 с.
7. Навчання нейромережі з учителем, без вчителя, з підкріпленням – у чому відмінність? URL: <https://neurohive.io/ru/osnovy-data-science/obuchenie-s-uchitelem-bez-uchitelja-s-podkrepleniem/>(дата звернення: 15.10.2022).
8. Плєскач В.Л., Рогушина Ю.В., Кустова Н.П. Інформаційні технології та системи. К.: Книга, 2004. 519 с.

9. Постанова КМУ «Про затвердження Примірного енергосервісного договору» від 21.10.2015 №845 [Електронний ресурс]. Верховна Рада України. Режим доступу : <http://zakon5.rada.gov.ua/laws/show/845-2015-п>.

10. Проект Закону України «Про ринок електричної енергії України» [Електронний ресурс]. Міністерство енергетики та вугільної промисловості України. Режим доступу: <http://mpe.kmu.gov.ua/minugol/doccatalog/document?id=245044194>.

11. Ропорядження КМУ «Про затвердження Національного плану дій з енергоефективності на період до 2020 р.» від 25.11.2015 № 1228-р [Електронний ресурс]. Верховна Рада України. Режим доступу: <http://zakon3.rada.gov.ua/laws/file/text/40/f450549n7.zip>

12. Чумаченко, Е.И. Алгоритм решения задачи прогнозирования / Е.И. Чумаченко, В.С. Горбатюк // Искусственный интеллект. – 2012. – № 2. – С. 24–30.

13. Ahmed A.M., Rizaner A., and Ulusoy A.H., A Decision Tree Algorithm Combined with Linear Regression for Data Classification, 2018 Int. Conf. Comput. Control. Electr. Electron. Eng. ICCCEEE 2018, no. August, pp. 1–5, doi: 10.1109/ICCCEEE.2018.8515759.

14. Bevans R. Simple Linear Regression: An Easy Introduction & Examples. 2020. URL: <https://www.scribbr.com/statistics/multiple-linear-regression/>

15. Breiman L., Friedman J., Charles J. Stone, OlshenTaylor & Francis R.A. Classification and Regression Trees. 1984. P. 368.

16. Catalina T., Virgone J., and Blanco E., Development and validation of regression models to predict monthly heating demand for residential buildings. Energy Build. ENERG BLDG, vol. 40, pp. 1825–1832, Dec. 2008, doi: 10.1016/j.enbuild.2008.04.001.

17. Dong B., Cao C., and Lee S. E. Applying support vector machines to predict building energy consumption in tropical region. Scopus, May 2005, Accessed: Nov. 30, 2019. [Online]. Available: <https://scholarbank.nus.edu.sg/handle/10635/113973>.

18. Ebrahimi N. Modeling, Simulation and Control of a Robotic Arm. engrXiv, preprint, Aug. 2019. doi: 10.31224/osf.io/t8fsr.
19. Elman J.L. Finding structure in time. *Cognitive Science*. 1990. 14. P. 179–211.
20. Energy efficiency Data Set. URL: <https://archive.ics.uci.edu/ml/datasets/energy+efficiency>
21. European Commission, “Directive 2002/91/EC of the European Parliament and of the Council of 16th December 2002 on the energy performance of buildings,” *Off. J. Eur. Communities*, vol. L1/65 – L1/71, Apr. 2003.
22. Hertzmann A., Fleet D. J., and Brubaker M. *Linear Regression*. 2015. URL: <http://www.cs.toronto.edu/~mbrubake/teaching/C11/Handouts/LinearRegression.pdf>
23. Hoshyarmanesh H., Ebrahimi N., Jafari A., Hoshyarmanesh P., Kim M., and Park H.-H. PZT/PZT and PZT/BiT Composite Piezo-Sensors in Aerospace SHM Applications: Photochemical Metal Organic + Infiltration Deposition and Characterization. *Sensors*, vol. 19, no. 1, p. 13, Jan. 2019, doi: 10.3390/s19010013.
24. Jordan M. Constrained supervised learning. *Journal of Mathematical Psychology*. 1992. 36. P. 396–452.
25. Mandic D.P., Chambers J.A. *Recurrent Neural Networks for Prediction*. Chichester: John Wiley&Sons, 2001. 285 p.
26. Moeedlodhi. How Outliers Can Pose a Problem in Linear Regression. URL: <https://medium.com/swlh/how-outlierscan-pose-a-problem-in-linear-regression-1431c50a8e0>
27. Molnar C. *Interpretable machine learning. A Guide for Making Black Box Models Explainable*, 2019. URL: <https://christophm.github.io/interpretable-ml-book/>.
28. Tibshirani S. and Friedman H. Random Forest Regression model explained in depth, GDCoder, Jun. 04, 2019. <https://gdccoder.com/random-forest->

regressor-explained-in-depth/ (accessed Nov. 30, 2019). Valerie and Patrick Hastie, p. 764.

29. Tsanas A. and Xifara A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy Build.* vol. 49, pp. 560–567, Jun. 2012, doi: 10.1016/j.enbuild.2012.03.003.

30. Tzafestas S., Tzafestas S. Computational intelligence techniques for short-term electric load forecasting. *Journal of Intelligent and Robotic Systems.* 2001. 31. P. 7–68.

31. UCI Machine Learning Repository: Energy efficiency Data Set. <http://archive.ics.uci.edu/ml/datasets/Energy+efficiency> (accessed Nov. 30, 2019).

32. Vikko N., Lautala P. Short-term electric power production scheduling using simulated annealing algorithm: Proc. of the IASTED Inter. Sym. ACTA Press, Anaheim, CA, USA, 1990.

33. Williams R.J., D. A. Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation.* 1989. 1. P. 270–280.

34. Elman J.L. Finding structure in time. *Cognitive Science.* 1990. 14. P. 179–211.

35. Williams R.J., Zipser D. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation.* 1989. 1. P. 270–280.

36. Yezioro A., Dong B., and Leite F. An applied artificial intelligence approach towards assessing building performance simulation tools. *Energy Build.* vol. 40, no. 4, pp. 612–620, Jan. 2008, doi: 10.1016/j.enbuild.2007.04.014.

37. Yu Z., Haghghat F., Fung B. C. M., and Yoshino H., A decision tree method for building energy demand modeling. *Energy Build.* vol. 42, no. 10, pp. 1637–1646, Oct. 2010, doi: 10.1016/j.enbuild.2010.04.006.

38. Zhang J. and Haghghat F., Development of Artificial Neural Network based heat convection algorithm for thermal simulation of large rectangular cross-sectional area Earth-to-Air Heat Exchangers. *Energy Build.* vol. 42, pp. 435– 440, Apr. 2010, doi: 10.1016/j.enbuild.2009.10.011.

39. Марк Саммерфилд. Программирование на Python 3. Подробное руководство. Пер. с англ. СПб.: Символ-Плюс, 2013. 608 с.
40. Лутц М. Программирование на Python. I том. СПб.: Символ-плюс, 2015. 992 с.
41. Прохоренок Н.А. Python 3 и PyQt. Разработка приложений. СПб.: БХВ-Петербург, 2012. 704 с.
42. Хахаев И.А. Практикум по алгоритмизации и программированию на Python. М.: Альт Линукс, 2010. 126 с.
43. Downey A., Elkner J., Meyers Ch. How to Think Like a Computer Scientist: Learning with Python. - Wellesley, Massachusetts: Green Tea Press, 2002. 290 pp.
44. Tutorialspoint / Python [Электронный ресурс]. Режим доступа: <http://www.tutorialspoint.com/python/>