

Міністерство освіти і науки України
Університет митної справи та фінансів

Факультет інноваційних технологій
Кафедра комп'ютерних наук та інженерії програмного забезпечення

Кваліфікаційна робота магістра

на тему: «Застосування машинного навчання для аналізу даних споживання електроенергії»

Виконав: студент групи K23-2M

Спеціальність 122 Комп'ютерні науки

Звоненко Б.Є.

(прізвище та ініціали)

Керівник д.т.н., проф. Яковенко В. О.

(науковий ступінь, вчене звання, прізвище та ініціали)

Рецензент Дніпровський державний

технічний університет

(місце роботи)

доцент кафедри математичного

моделювання та системного аналізу

(посада)

к.т.н., доц. Волосова Н.М.

(науковий ступінь, вчене звання, прізвище та ініціали)

АНОТАЦІЯ

Звоненко Б.Є. Застосування машинного навчання для аналізу даних споживання електроенергії.

Дипломна робота на здобуття освітнього ступеня магістр за спеціальністю 122 «Комп'ютерні науки» – Університет митної справи та фінансів, Дніпро, 2025.

Кваліфікаційна робота магістра присвячена дослідженню застосування методів машинного навчання для аналізу та прогнозування даних споживання електроенергії. Машинне навчання, як один із найперспективніших напрямків сучасної науки, надає нові можливості для виявлення закономірностей у великих обсягах даних, створення точних моделей прогнозування, оптимізації використання ресурсів та підвищення стабільності енергетичних мереж.

У роботі проведено детальний аналіз предметної області, охарактеризовано сучасні методи обробки та прогнозування даних у сфері енергетики. Зокрема, розглянуто статистичні підходи, методи регресії, класифікації, кластеризації та алгоритми глибокого навчання. Особливу увагу приділено нейронним мережам, таким як рекурентні моделі та LSTM, які здатні ефективно працювати з часовими рядами, враховуючи довгострокові залежності та вплив зовнішніх факторів, таких як погодні умови, сезонність і соціально-економічні зміни.

Практична значимість роботи полягає в розробці і впровадженні систем прогнозування та аналізу, які дозволяють енергетичним компаніям оптимізувати витрати, підвищити ефективність розподілу ресурсів, попереджати аварійні ситуації та сприяти екологічній стійкості.

Ключові слова: машинне навчання, прогнозування споживання електроенергії, нейронні мережі, обробка даних, енергетичні системи, оптимізація ресурсів, аналіз часових рядів, виявлення аномалій.

ABSTRACT

Zvonenko B.Ye. Application of machine learning to analyze electricity consumption data.

Master's thesis for the degree of Master's Degree in Computer Science (specialty 122) – University of Customs and Finance, Dnipro, 2025.

The master's thesis is devoted to the study of the application of machine learning methods for analyzing and forecasting electricity consumption data. Machine learning, as one of the most promising areas of modern science, provides new opportunities for identifying patterns in large amounts of data, creating accurate forecasting models, optimizing resource use, and increasing the stability of energy networks.

The paper provides a detailed analysis of the subject area, characterizes modern methods of data processing and forecasting in the energy sector. In particular, statistical approaches, regression, classification, clustering, and deep learning algorithms are considered. Particular attention is paid to neural networks, such as recurrent models and LSTMs, which are able to work effectively with time series, taking into account long-term dependencies and the influence of external factors such as weather conditions, seasonality and socio-economic changes.

The practical significance of the work lies in the development and implementation of forecasting and analysis systems that allow energy companies to optimize costs, increase the efficiency of resource allocation, prevent emergencies, and promote environmental sustainability.

Keywords: machine learning, electricity consumption forecasting, neural networks, data processing, energy systems, resource optimization, time series analysis, anomaly detection.

ЗМІСТ

ВСТУП	6
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ	10
1.1 Класифікація методів аналізу енергетичних даних.....	10
1.2 Класичні підходи до прогнозування споживання електроенергії	13
1.3 Застосування статистичних методів у прогнозуванні споживання електроенергії.....	16
1.4 Основні підходи машинного навчання в енергетичній галузі	20
1.5 Аналіз досліджень.....	24
1.6 Висновки до першого розділу	32
РОЗДІЛ 2. ДОСЛІДЖЕННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ СТВОРЕННЯ МОДЕЛІ АНАЛІЗУ ДАНИХ СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ.....	35
2.1 Техніки навчання в енергетичних даних.....	35
2.2 Використання нейронних мереж у прогнозуванні споживання електроенергії.....	39
2.3 Методи очищення та підготовки даних для аналізу споживання електроенергії.....	43
2.4 Проблеми та особливості роботи з великими даними в енергетичній сфері	46
2.5 Аналіз залежності споживання електроенергії від зовнішніх факторів	50
2.6 Вибір методу навчання.....	53
2.7 Висновки до другого розділу.....	56
РОЗДІЛ 3. ПРОГРАМНА РЕАЛІЗАЦІЯ.....	58
3.1 Мета програмної реалізації.....	58
3.2 Інструменти розробки.....	59
3.3 Опис реалізації	61

3.4 Навчання моделі.....	64
3.5 Тестування моделі.....	67
3.6 Можливість інтеграції з іншими інструментами.....	73
3.7 Напрямки подальших досліджень.....	76
3.8 Висновки до третього розділу	79
ВИСНОВКИ.....	80
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	82
ДОДАТКИ.....	85

ВСТУП

У сучасному світі, де інформаційні технології та автоматизація стають основою розвитку різних галузей економіки, важливу роль відіграє ефективне управління енергетичними ресурсами. Одним із найбільш важливих аспектів у цій сфері є аналіз і оптимізація споживання електроенергії. Це завдання має не лише економічну, а й екологічну значущість, адже в умовах зростаючого попиту на енергоресурси та необхідності зменшення викидів парникових газів, ефективне використання енергії є важливим кроком до стійкого розвитку. Одним із потужних інструментів для аналізу даних, які стосуються споживання електроенергії, є методи машинного навчання. Вони дозволяють виявляти складні закономірності у великих обсягах даних, що є необхідним для створення ефективних систем моніторингу та прогнозування потреби в електроенергії. Зокрема, застосування машинного навчання в аналізі даних споживання електроенергії відкриває нові можливості для точного прогнозування навантажень на енергетичні мережі, виявлення аномалій у споживанні та розробки рекомендацій для оптимізації енергетичних витрат. Це, в свою чергу, може значно знизити енергетичні витрати, підвищити ефективність роботи енергетичних підприємств та сприяти сталому розвитку енергетичних систем.

Актуальність дослідження застосування машинного навчання для аналізу даних споживання електроенергії обумовлена декількома чинниками. З одного боку, зростаюче споживання енергоресурсів вимагає розвитку нових технологій та методів для точного прогнозування потреб електроенергії і забезпечення її раціонального використання. З іншого боку, машинне навчання є одним із найперспективніших напрямків для вирішення цієї проблеми, оскільки дозволяє автоматизувати процеси збору та аналізу даних, а також створювати прогностичні моделі, які здатні враховувати різноманітні фактори впливу.

Машинне навчання, завдяки своїм можливостям працювати з великими обсягами даних та адаптуватися до змінюваних умов, дозволяє не лише покращити точність прогнозів, але й виявляти аномальні ситуації в режимі реального часу. Це дозволяє оперативно реагувати на можливі неполадки в енергетичних мережах, знижувати витрати енергії та оптимізувати роботу енергосистем, що є особливо важливим у контексті глобальних тенденцій збереження енергоресурсів і боротьби зі змінами клімату.

Метою даного дослідження є розробка і впровадження методів машинного навчання для аналізу та прогнозування даних споживання електроенергії, що дозволяють підвищити ефективність управління енергетичними системами. Для досягнення цієї мети необхідно вирішити низку завдань:

- оцінити поточний стан методів машинного навчання в контексті аналізу енергетичних даних, зокрема, визначити найбільш ефективні підходи до обробки великих обсягів інформації;

- розробити математичні моделі для прогнозування споживання електроенергії на основі даних про попит, кліматичні умови, соціально-економічні фактори та інші параметри;

- розробити алгоритми машинного навчання для виявлення аномалій у споживанні електроенергії та прогнозування можливих відхилень;

- оцінити ефективність запропонованих моделей шляхом порівняння з існуючими методами прогнозування та аналізу.

Об'єктом дослідження є процеси аналізу і прогнозування споживання електроенергії на основі даних, що зберігаються в енергетичних системах та базах даних. Предметом дослідження є методи машинного навчання, зокрема алгоритми регресії, класифікації, а також глибинного навчання, які можуть бути використані для моделювання споживання енергії та виявлення аномальних ситуацій.

У дослідженні буде використано комплексний підхід, що включає теоретичні та практичні методи. До теоретичних методів належать:

- аналіз наукових праць та літератури, що стосується машинного навчання та енергетичних систем;
- математичне моделювання процесів споживання електроенергії;
- методи обробки та аналізу великих даних.

Практичні методи дослідження включають:

- розробку та апробацію алгоритмів машинного навчання для прогнозування та аналізу даних;
- тестування моделей на реальних даних про споживання електроенергії;
- аналіз точності та ефективності запропонованих моделей за допомогою порівняння з існуючими методами.

Практична значимість роботи полягає у створенні нових підходів для ефективного управління енергетичними системами за допомогою передових технологій машинного навчання. Результати дослідження можуть бути використані для оптимізації процесів прогнозування споживання електроенергії, що дозволить знизити витрати енергії, покращити стабільність енергетичних мереж та зменшити екологічний вплив енергетичних процесів.

Також запропоновані методи можуть бути застосовані для виявлення аномалій в споживанні енергії, що дозволить оперативно реагувати на несанкціоноване споживання або неполадки в мережах, що сприятиме підвищенню рівня безпеки енергетичних систем.

Наукова новизна роботи полягає у впровадженні нових методів машинного навчання для аналізу та прогнозування споживання електроенергії, зокрема в контексті створення комбінованих моделей, що враховують різні фактори, які можуть впливати на споживання енергії. Крім того, дослідження сприятиме розвитку теоретичних аспектів застосування машинного навчання для розв'язання задач енергетики, що є актуальним в

умовах розвитку «розумних» енергетичних мереж та підвищення вимог до точності та швидкості прогнозів.

Кваліфікаційна робота сприятиме як розвитку науки в області енергетики та машинного навчання, так і наданню практичних рішень для реальних проблем енергетичних компаній.

Структура кваліфікаційної роботи магістра. Кваліфікаційна робота складається з трьох розділів. Обсяг кваліфікаційної роботи – 94 сторінки. Робота містить 11 рисунків. Список використаних джерел має 14 посилань.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

1.1 Класифікація методів аналізу енергетичних даних

Класифікація методів аналізу енергетичних даних є важливою складовою частиною сучасної енергетики, оскільки дозволяє ефективно обробляти великі обсяги інформації, що отримуються від різних енергетичних систем, таких як електричні, теплові, гідравлічні та інші (рис. 1.1). Враховуючи постійне зростання потреби у високоякісних та надійних методах оцінки енергетичних процесів, вчені та інженери розробляють численні підходи, які дозволяють оптимізувати експлуатацію енергетичних систем, знижувати витрати енергії, підвищувати ефективність використання ресурсів і зменшувати негативний вплив на навколишнє середовище [1].

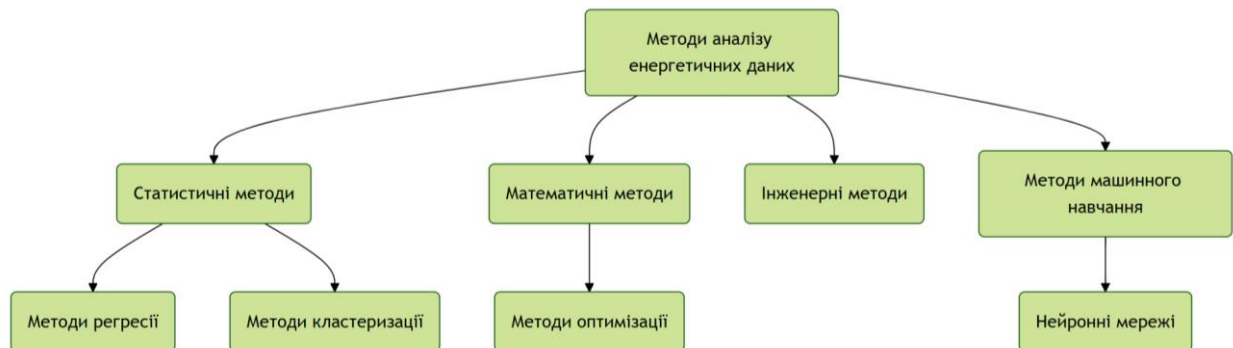


Рисунок 1.1 – Класифікація методів аналізу енергетичних даних

Методи аналізу енергетичних даних поділяються на кілька основних категорій, зокрема, статистичні, математичні, інженерні, а також методи на основі машинного навчання та штучного інтелекту. Кожен з цих підходів має свої особливості, які дозволяють застосовувати їх до різних типів енергетичних систем і задач. Статистичні методи аналізу енергетичних даних

передбачають використання різноманітних статистичних інструментів для виявлення закономірностей і тенденцій у великих масивах даних. Одним з таких методів є кореляційний аналіз, який дозволяє виявити взаємозв'язки між різними змінними, що характеризують енергетичні процеси. Для цього використовуються різні кореляційні показники, такі як коефіцієнт кореляції Пірсона, який вимірює силу і напрямок лінійного зв'язку між двома змінними. Іншим важливим статистичним методом є регресійний аналіз, що дозволяє побудувати модель, яка описує залежність однієї змінної від інших і передбачає можливі зміни в енергетичних параметрах в залежності від зовнішніх умов.

Застосування статистичних методів у енергетичних системах є особливо корисним при аналізі часових рядів даних, оскільки багато енергетичних параметрів, таких як споживання енергії, температура, вологість тощо, змінюються в часі. Для таких цілей широко використовуються методи аналізу часових рядів, зокрема методи ковзних середніх, авто- та кроскореляційний аналіз, а також методи прогнозування, такі як авторегресивні інтегровані середні (ARIMA) моделі, які дозволяють робити прогнози на основі історичних даних.

Математичні методи аналізу енергетичних даних використовуються для побудови моделей, що описують енергетичні процеси в різних системах. Одним з таких методів є методи оптимізації, які дозволяють знаходити найкращі рішення для управління енергетичними системами, зокрема для мінімізації витрат енергії або максимізації її ефективності [2]. Для цього використовуються різні підходи, такі як лінійне та нелінійне програмування, методи динамічного програмування, а також методи стійкості, які дозволяють оцінити здатність енергетичної системи до підтримки стабільності за різних умов експлуатації. Одним з ключових напрямів математичного аналізу енергетичних даних є моделювання енергетичних процесів за допомогою диференціальних рівнянь, які описують зміни енергетичних параметрів у часі.

Такі моделі можуть бути використані для аналізу термодинамічних процесів в електростанціях, теплових мережах, а також для прогнозування потреби в енергії в певних регіонах. Система таких рівнянь часто включає в себе параметри, що описують потоки енергії, теплообмін, ефективність роботи обладнання та інші важливі фактори, які визначають поведінку енергетичної системи в різних умовах.

Інженерні методи аналізу енергетичних даних фокусуються на практичному застосуванні теоретичних знань для оптимізації роботи енергетичних систем. Це включає в себе застосування моделей для проектування та налаштування енергетичних установок, а також для моніторингу та управління їхньою експлуатацією. Одним з важливих аспектів є використання методів діагностики та прогнозування для попередження аварійних ситуацій та оптимізації технічного обслуговування енергетичних систем. Для цього використовуються такі інженерні інструменти, як аналіз надійності, методи оцінки ефективності, а також методи оптимізації параметрів роботи устаткування.

Сучасні підходи до аналізу енергетичних даних включають використання методів на основі машинного навчання та штучного інтелекту, які дозволяють значно підвищити ефективність обробки великих обсягів даних та покращити точність прогнозування енергетичних процесів. Машинне навчання дозволяє автоматично виявляти складні залежності між параметрами енергетичних систем, які можуть бути неочевидними за допомогою традиційних статистичних чи математичних методів. Одним з найбільш поширених методів є використання нейронних мереж, які здатні моделювати складні нелінійні залежності і здійснювати прогнози на основі великих масивів даних. Крім того, використовуються методи глибинного навчання, що дозволяють обробляти ще більші обсяги даних і досягати високої точності при прогнозуванні енергетичних параметрів.

Також велике значення для аналізу енергетичних даних мають методи кластеризації, які дозволяють групувати схожі за характеристиками енергетичні об'єкти чи процеси [1]. Це дає змогу виявляти аномалії в роботі систем, знижувати ризики непередбачуваних ситуацій і оптимізувати процеси управління енергетичними ресурсами. За допомогою таких методів можна також здійснювати сегментацію споживачів енергії, що дає змогу створювати персоналізовані стратегії для зниження споживання та поліпшення ефективності.

Таким чином, незважаючи на різноманіття підходів, усі методи аналізу енергетичних даних мають спільну мету – підвищити ефективність управління енергетичними системами та забезпечити сталий розвиток енергетичних процесів. Вибір конкретного методу залежить від специфіки задачі, обсягу та типу даних, доступних для аналізу, а також від необхідної точності результатів. Враховуючи швидкий розвиток технологій, застосування новітніх методів обробки енергетичних даних стає важливим елементом інноваційного розвитку енергетичних систем на глобальному рівні.

1.2 Класичні підходи до прогнозування споживання електроенергії

Прогнозування споживання електроенергії є важливим аспектом управління енергетичними системами та планування їхнього розвитку. В умовах зростаючого попиту на електроенергію та необхідності забезпечення стабільного постачання енергоресурсів, точне передбачення потреби в електроенергії допомагає не лише в ефективному управлінні енергетичними мережами, але й у забезпеченні економічної ефективності роботи енергетичних підприємств. Класичні підходи до прогнозування споживання електроенергії, що використовуються в сучасній енергетиці, відрізняються різноманіттям методів, кожен з яких має свої переваги та обмеження [2]. Усі ці підходи ґрунтуються на аналізі історичних даних і застосуванні

математичних моделей для визначення закономірностей у зміні споживання енергоресурсів у часі. Одним з основних напрямків у класичних методах прогнозування є використання статистичних моделей. Статистичні методи прогнозування споживання електроенергії здебільшого базуються на аналізі часових рядів, тобто на даних про споживання електроенергії, зафіксованих протягом певного періоду часу. Найпоширенішими серед них є методи лінійної та нелінійної регресії, а також методи, що використовують авто- та кроскореляційний аналіз. Регресійний аналіз дозволяє побудувати модель, яка відображає залежність між рівнем споживання електроенергії та різними факторами, такими як температура повітря, день тижня, сезонність, економічні показники, демографічні зміни та інші. Оскільки електричне споживання зазвичай є функцією від багатьох змінних, застосування таких моделей допомагає врахувати ці взаємозв'язки та здійснити більш точні прогнози.

Лінійна регресія є одним з найпростіших методів, де залежність між споживанням електроенергії та незалежними змінними (наприклад, температурою або числом робочих днів) представлена лінійною функцією. Однак, такий підхід має суттєві обмеження, оскільки багато енергетичних процесів є нелінійними. Нелінійна регресія дозволяє побудувати більш складні моделі, які можуть краще відображати реальні процеси, що відбуваються в енергетичних системах, але водночас ці моделі є більш чутливими до вхідних даних і вимагають більш точного налаштування [1].

Автокореляційні методи, зокрема авторегресивні інтегровані середні моделі (ARIMA), також широко використовуються для прогнозування споживання електроенергії. Основною особливістю цих методів є те, що вони ґрунтуються на аналізі часових рядів даних і використовують інформацію про попередні значення величини для передбачення майбутніх. ARIMA-моделі дозволяють здійснювати прогнозування навіть за наявності сезонних коливань у споживанні електроенергії, таких як підвищене споживання в зимовий період через обігрівання приміщень або влітку через охолодження. Вони

враховують як рівень минулого споживання, так і вплив різних зовнішніх факторів на енергетичні потреби.

Іншим важливим методом є методи, що використовують модель експоненційного згладжування, які є особливо корисними при наявності змінної сезонності або трендів у споживанні. Експоненційне згладжування дозволяє надавати різну вагу останнім спостереженням, що робить його ефективним інструментом для прогнозування короткострокових змін в енергоспоживанні. Ці методи передбачають, що майбутні значення споживання енергії можна оцінити як комбінацію минулих значень з врахуванням певного коефіцієнта згладжування.

Водночас, класичні методи прогнозування, зокрема статистичні та математичні моделі, мають свої обмеження. Вони часто не враховують складні нелінійні взаємодії між різними факторами, які впливають на споживання електроенергії, і можуть демонструвати незначну точність при прогнозуванні у випадках, коли ці фактори мають сильний вплив. Також, велику роль відіграє точність і повнота вхідних даних, оскільки невірні або неповні дані можуть призвести до значних помилок у прогнозах [3].

Незважаючи на це, статистичні та математичні методи прогнозування споживання електроенергії продовжують залишатися основними інструментами для планування енергетичних ресурсів, оскільки вони здатні забезпечити високий рівень точності при наявності великої кількості даних і стабільних умов експлуатації енергетичних систем. Важливим аспектом є також можливість комбінувати різні методи, використовуючи їхні переваги для досягнення кращих результатів. Наприклад, можна поєднувати методи регресії з авторегресійними моделями або застосовувати екстраполяцію трендів до результатів, отриманих за допомогою статистичних моделей.

Окрім чисто математичних та статистичних підходів, класичні методи прогнозування включають також інженерні підходи, які ґрунтуються на глибокому розумінні специфіки роботи енергетичних систем і фізичних

процесів, що в них відбуваються. Наприклад, для прогнозування споживання електроенергії можна використовувати моделі, що описують енергетичні потреби з урахуванням характеристик обладнання, кліматичних умов, а також соціально-економічних аспектів, таких як зміни в рівні життя, урбанізація, зміни в споживчих звичках тощо. Ці моделі можуть бути особливо корисними для довгострокового прогнозування, коли важливо враховувати не тільки поточні тенденції, але й майбутні зміни в структурі споживання енергії.

Таким чином, класичні підходи до прогнозування споживання електроенергії є важливою складовою частиною сучасної енергетики і дають змогу забезпечити ефективне планування енергетичних ресурсів. Однак, для досягнення більш точних результатів необхідно враховувати різноманітні фактори, що впливають на споживання електроенергії, і використовувати комбіновані методи, які здатні компенсувати недоліки кожного з підходів. Тільки так можна отримати точні прогнози, які допоможуть забезпечити стабільну і ефективну роботу енергетичних систем у різних умовах.

1.3 Застосування статистичних методів у прогнозуванні споживання електроенергії

Прогнозування споживання електроенергії є важливою складовою частиною енергетичного планування та управління енергетичними ресурсами. Для досягнення високої точності в прогнозах необхідно використовувати науково обґрунтовані підходи, які враховують не тільки фактори, що безпосередньо впливають на споживання, але й вивчають взаємозв'язки між ними [2, 3]. Одним із таких підходів є застосування статистичних методів, які є важливими інструментами для аналізу та прогнозування змін у споживанні електроенергії. Статистичні методи дозволяють систематизувати велику кількість даних, виявляти приховані закономірності та тренди, що дозволяє забезпечити більш точні прогнози споживання електричної енергії.

Перелік класичних методів прогнозування наведено на рисунку 1.2.

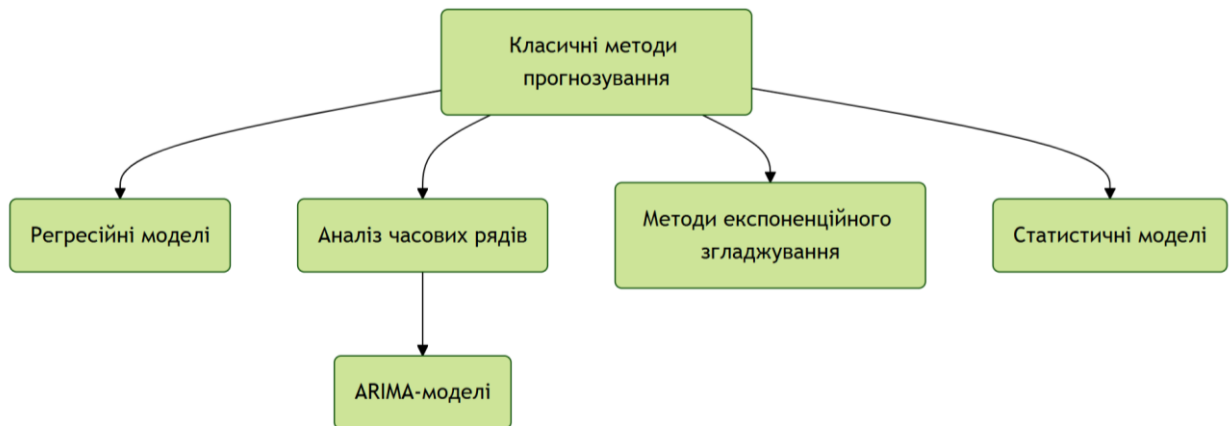


Рисунок 1.2 – Класичні методи прогнозування

Оскільки споживання електроенергії є функцією від багатьох факторів, серед яких погодні умови, добові та сезонні коливання, економічні показники, зміни в інфраструктурі та соціально-демографічні тенденції, статистичні методи дозволяють обробляти великі обсяги даних і здійснювати їх аналіз. Використання таких методів дозволяє отримати змодельовані залежності між різними параметрами, що допомагають виявити ключові фактори, які визначають зміни в споживанні енергії.

Один із найбільш поширених статистичних методів, що застосовуються у прогнозуванні споживання електроенергії, – це кореляційний аналіз. Цей метод дозволяє виявити взаємозв'язки між різними змінними, що впливають на споживання електроенергії, такими як температура повітря, рівень вологості, кількість сонячних годин, сезонність, час доби, кількість населення в регіоні та інші фактори. Оскільки споживання енергії в значній мірі залежить від погодних умов, кореляційний аналіз дозволяє оцінити, як саме температура або інші атмосферні умови впливають на електричне споживання. Це важливо, оскільки дозволяє створювати прогнози для конкретних періодів року, коли енергетичне навантаження є особливо великим або малим [3]. Також одним із

широко використовуваних методів є регресійний аналіз. Зазвичай цей метод полягає в побудові математичної моделі, що описує залежність між споживанням електроенергії та іншими факторами, наприклад, температурою, рівнем вологості, попитом на промислову продукцію або змінами в рівні споживання серед населення. Регресія дозволяє оцінити величину впливу кожного з цих факторів на електричне споживання та використати отримані дані для прогнозування майбутніх змін у попиті на електроенергію. Лінійна регресія є базовим методом, коли залежність між змінними можна виразити прямою лінією. Проте в реальних умовах споживання електроенергії може змінюватися нелінійно, тому для більш складних ситуацій використовуються методи нелінійної регресії.

Існують також методи, які використовують моделі часових рядів для прогнозування споживання електроенергії. Часові ряди є одним із найбільш важливих інструментів при прогнозуванні в енергетиці, оскільки багато енергетичних процесів є залежними від часу. Вони відображають зміну споживання в часі, дозволяючи виявити повторювані тренди, сезонні коливання або циклічні зміни. Для аналізу таких даних використовуються методи, які можуть допомогти у визначенні залежностей між поточними значеннями споживання та попередніми спостереженнями. Одним із таких методів є авторегресивні моделі, такі як ARIMA (авторегресивні інтегровані середні). Ці моделі дозволяють здійснити прогнозування на основі історичних даних, при цьому вони здатні враховувати як сезонні коливання, так і інші фактори, що мають довготривалі впливи на споживання енергії.

ARIMA є одним із найбільш популярних методів прогнозування, оскільки його основна перевага полягає в здатності працювати навіть за наявності складних змін у структурі даних. Моделі ARIMA можуть враховувати як прямі залежності, так і залежності з лагами, тобто затримками в часі. Це дозволяє робити більш точні прогнози, які є важливими при плануванні споживання електроенергії в довгостроковій перспективі, оскільки

енергетичні коливання можуть мати різні цикли, що не завжди очевидні на перший погляд [2-4]. Іншим важливим методом є експоненційне згладжування, яке широко застосовується в практиці прогнозування енергоспоживання. Цей метод полягає в тому, що останнім спостереженням надається більше значення при обчисленні прогнозу, що дозволяє враховувати швидкі зміни в споживанні. Експоненціальне згладжування є корисним для короткострокових прогнозів, коли зміни в споживанні відбуваються швидко, а дані про поточне споживання мають важливе значення для передбачення наступних періодів. Наприклад, для прогнозування споживання в межах доби або тижня цей метод може бути дуже корисним, оскільки він дозволяє швидко реагувати на зміни, пов'язані з погодними умовами або різкими змінами в економічній активності.

Незважаючи на свою ефективність, статистичні методи мають ряд обмежень, які можуть вплинути на точність прогнозів. Одним з основних недоліків є необхідність наявності великих обсягів історичних даних, оскільки точність моделей безпосередньо залежить від кількості доступних спостережень. Крім того, статистичні методи часто не можуть адекватно враховувати всі можливі фактори, що впливають на споживання електроенергії. Наприклад, вони можуть бути не в змозі повністю врахувати випадкові або надзвичайні ситуації, такі як стихійні лиха або раптові економічні кризи, які можуть різко змінити рівень попиту на електроенергію.

Тому на додаток до класичних статистичних методів, для прогнозування споживання електроенергії часто використовуються більш складні підходи, такі як методи машинного навчання або глибоке навчання, які здатні враховувати більш складні нелінійні залежності та взаємодії між великою кількістю змінних. Проте статистичні методи залишаються важливим інструментом для аналізу даних, оскільки вони дозволяють здійснити попередній аналіз, побудувати базові моделі та виявити основні тренди в змінах споживання енергії [4].

Таким чином, застосування статистичних методів у прогнозуванні споживання електроенергії є важливим елементом у побудові точних і надійних моделей, які допомагають у прийнятті рішень щодо управління енергетичними системами. Вони дозволяють зменшити невизначеність в прогнозах, що є критичним для ефективного планування енергетичних ресурсів, а також для забезпечення стабільного функціонування енергетичних мереж і зниження витрат на виробництво та транспортування електроенергії.

1.4 Основні підходи машинного навчання в енергетичній галузі

Машинне навчання в енергетичній галузі є одним з найперспективніших напрямків розвитку сучасних технологій, що дозволяє значно підвищити ефективність енергетичних систем, оптимізувати використання ресурсів, покращити управління споживанням енергії та знизити витрати. У зв'язку з розвитком великих даних, потужних обчислювальних систем і нових алгоритмів машинного навчання, енергетична галузь переживає етап значних змін, що відкривають нові можливості для більш точного прогнозування, автоматизації процесів та інтелектуального управління енергетичними системами [3, 4]. Машинне навчання є потужним інструментом для вирішення складних завдань, таких як оптимізація споживання електричної енергії, управління розподілом навантажень, забезпечення стабільності енергетичних мереж, а також для ефективного прогнозування та керування відновлюваними джерелами енергії, що характеризуються високою мінливістю.

В основі машинного навчання лежить використання алгоритмів, що дозволяють системам навчатися на даних, виявляти закономірності та робити прогнози або класифікацію без необхідності явно заданих правил. Завдяки цим властивостям машинне навчання відкриває нові горизонти для використання у складних енергетичних задачах, що включають не тільки прогнозування, а й адаптивне управління та підтримку прийняття рішень у реальному часі.

Оскільки енергетичні системи складаються з численних складних компонентів і параметрів, де часто спостерігаються нелінійні взаємодії, алгоритми машинного навчання стають незамінними для вирішення таких завдань, як ідентифікація патернів у споживанні енергії, оптимізація параметрів енергетичних систем або моніторинг та діагностика технічних систем.

Основними напрямками застосування машинного навчання в енергетичній галузі є прогнозування споживання енергії, прогнозування та оптимізація роботи відновлюваних джерел енергії, виявлення аномалій і діагностика несправностей у мережах, управління розподіленими енергетичними ресурсами та побудова інтелектуальних систем для управління енергоспоживанням у побутових і промислових умовах. Алгоритми машинного навчання використовуються для побудови математичних моделей, які здатні враховувати не тільки традиційні фактори, такі як температура, вологість або час доби, але й більш складні взаємозв'язки між численними змінними, що впливають на споживання енергії, що дозволяє робити прогнози більш точними [3-5]. Одним з основних підходів машинного навчання в енергетичній галузі є використання методів регресії для прогнозування величини споживання електричної енергії. Оскільки споживання енергії є складною функцією від багатьох параметрів, таких як температура, вологість, економічна активність, зміни в популяції, наявність святкових днів або інші соціальні фактори, застосування регресійних моделей дозволяє побудувати математичні зв'язки між цими змінними та рівнем споживання енергії. Лінійна та нелінійна регресія, а також різні техніки машинного навчання, такі як методи опорних векторів (SVM), дерева рішень, випадкові ліси, або нейронні мережі, застосовуються для вирішення цієї задачі. Наприклад, нейронні мережі здатні враховувати складні нелінійні залежності та оптимізувати прогнози споживання енергії, забезпечуючи високу точність навіть у випадках, коли традиційні статистичні методи не дають задовільних результатів. За допомогою глибокого навчання та інших передових підходів

можна створювати більш складні моделі, які можуть враховувати сезонні коливання, добові цикли, а також інші фактори, що впливають на споживання енергії [4, 5]. Іншим важливим застосуванням машинного навчання в енергетичній сфері є прогнозування енергетичних потреб відновлюваних джерел енергії, таких як сонячні батареї або вітрові турбіни. Оскільки робота цих джерел енергії залежить від змінних факторів, таких як сонячна радіація або швидкість вітру, їхнє прогнозування вимагає використання складних математичних моделей, які можуть адаптуватися до змінних умов. У цьому контексті машинне навчання допомагає створювати моделі, які здатні не лише прогнозувати виробництво енергії в короткостроковій перспективі, але й оптимізувати процеси інтеграції відновлюваних джерел енергії в загальну енергетичну мережу. Це дозволяє значно підвищити ефективність використання таких джерел, зменшити залежність від традиційних енергетичних потужностей і забезпечити стабільне постачання енергії в мережу навіть за умов високої мінливості.

Машинне навчання також активно застосовується для виявлення аномалій і діагностики несправностей у енергетичних мережах. Мережі електропостачання, особливо в умовах високої складності, з великою кількістю взаємопов'язаних елементів, можуть зазнавати різних видів пошкоджень або перевантажень, які можуть вплинути на їхню роботу. Використання алгоритмів машинного навчання дозволяє вчасно виявляти аномальні ситуації, що можуть свідчити про можливі поломки або відхилення в роботі енергетичних систем, і, таким чином, забезпечити своєчасне реагування на проблеми. Зокрема, алгоритми на основі методів класифікації та аномалій, такі як методи опорних векторів (SVM) або глибокі нейронні мережі, можуть автоматично виявляти невідповідності в даних про стан мережі, виявляти проблеми і навіть передбачати можливі аварії або несправності. Це значно підвищує надійність енергетичних систем, оскільки

своєчасне виявлення аномалій дозволяє оперативно втручатися в процеси управління та запобігати великим аваріям [3].

Іншим важливим напрямком є управління розподіленими енергетичними ресурсами та інтелектуальні системи для управління енергоспоживанням у побутових і промислових умовах. Розвиток розумних мереж (smart grids) і інтелектуальних лічильників дозволяє зібрати великий обсяг даних про споживання електроенергії в реальному часі, що дає змогу застосовувати алгоритми машинного навчання для оптимізації процесів розподілу енергії, виявлення неефективних патернів споживання та надання рекомендацій щодо зниження витрат енергії. У таких системах машинне навчання використовується для аналізу поведінки користувачів, прогнозування споживання в різні періоди часу та адаптації енергетичних ресурсів до змінних умов попиту.

Завдяки інтелектуальним системам можна створити моделі, які не тільки прогнозують споживання енергії, а й оптимізують використання енергоресурсів на основі аналізу поведінки споживачів, сезонних та економічних факторів. Це дозволяє знижувати витрати на енергетичні ресурси та зменшувати навантаження на енергетичні мережі під час пікових навантажень, а також сприяти ефективному використанню відновлюваних джерел енергії. Алгоритми машинного навчання можуть бути використані для створення адаптивних систем, які самостійно коригують параметри енергоспоживання залежно від зміни умов.

Таким чином, застосування машинного навчання в енергетичній галузі є ключовим інструментом для підвищення ефективності та стійкості енергетичних систем. Завдяки здатності аналізувати великі обсяги даних, виявляти приховані закономірності та здійснювати адаптивне управління, машинне навчання дозволяє значно поліпшити процеси прогнозування, оптимізації та управління енергоспоживанням, а також забезпечити надійність і стабільність енергетичних мереж. Впровадження цих технологій у практику

енергетичних компаній і виробництво допоможе не тільки знизити витрати, а й підвищити ефективність використання енергетичних ресурсів у різних сферах економіки.

1.5 Аналіз досліджень

У статті [1] йдеться про використання алгоритму на основі Extreme Learning Machine (ELM) для аналізу поведінки споживачів електричної енергії в умовах розвитку інформаційних технологій та обробки даних. Зокрема, розглядається процес аналізу типових графіків навантаження, що допомагає у класифікації та аналізі споживчих даних. Для цього спочатку застосовують стратегію відбору ознак, щоб виділити найбільш релевантні характеристики графіків навантаження, що сприяє більш ефективному аналізу. Потім використовуються ці відібрані характеристики як вхід для алгоритму, а також проводиться оптимізація через вибір активаційних функцій та кількості нейронів у прихованому шарі для досягнення найкращих параметрів алгоритму ELM. Зрештою, результати порівнюються з методом нейронних мереж BP, і показано, що метод класифікації навантаження на основі характеристик поведінки споживання електричної енергії з використанням ELM дозволяє підвищити ефективність виявлення аномальних споживань та виявляти шаблони поведінки користувачів через аналіз даних споживання електроенергії.

У статті [2] розглядається метод передбачення споживання електричної енергії, що базується на використанні глибинного навчання. Зокрема, автори розробили систему для прогнозування споживання енергії в корпоративних будівлях, використовуючи набір даних, який містить дату, час та кількість спожитої енергії в мегаватах. Даний набір даних поділяється на дві частини: 80% для навчання моделі та 20% для тестування. Важливим аспектом дослідження є використання моделі глибинного навчання з архітектурою

LSTM (Long Short Term Memory) для прогнозування споживання енергії на різні періоди: годину, день, місяць чи рік. Для покращення ефективності моделі застосовується оптимізація за допомогою методу Adam. Модель була навчена на 50000 точках даних і протестована на 12874 точках. Прогнозоване споживання енергії також дозволяє обчислити можливий рахунок за спожиту електроенергію в рупіях. Результати експериментів показали, що алгоритм є точним і ефективним, що дає можливість точно прогнозувати споживання енергії та відповідні витрати. Додаткові переваги та деталі реалізації також обговорюються в статті.

Стаття [3] присвячена проблемі прогнозування споживання електричної енергії в домогосподарствах або будівлях, для якої було розроблено різні методи, такі як статистичні, машинного навчання та фізичні моделі, з метою покращення точності прогнозування. Це важлива задача як для споживачів, так і для постачальників електроенергії, оскільки точні прогнози дозволяють забезпечити баланс між постачанням та попитом на електроенергію, а також допомагають споживачам планувати витрати на електрику, особливо в умовах відключень або обмеженого доступу до енергетичних ресурсів. Стаття аналізує набір даних про споживання електроенергії одного домогосподарства і досліджує застосування методів статистики та машинного навчання для розробки моделей регресії для прогнозування споживання електричної енергії. Далі вивчається точність різних моделей регресії для прогнозування на 24 години вперед. Результати показують низьку точність прогнозування в усіх моделях, що виявляється в високих значеннях MAPE, близьких до 100%, що свідчить про великі помилки прогнозу, а також низькі значення R^2 , що означають, що моделі погано пояснюють варіації в цільовій змінній, що призводить до низької точності прогнозів.

Стаття [4] досліджує складні взаємозв'язки між споживанням енергії різними побутовими приладами та загальним споживанням електроенергії в домогосподарствах, звертаючись до методів машинного навчання. Основною

метою є розробка прогностичних моделей, які дозволяють оцінювати загальне споживання активної енергії, аналізуючи енергоспоживання в окремих приміщеннях, таких як кухня, пральня, система кондиціонування та електричний водонагрівач. Через ретельний аналіз і застосування методів машинного навчання стаття висвітлює основні зв'язки, що визначають патерни споживання енергії. Висновки цієї роботи мають важливе значення не лише з точки зору точності прогнозів, але й для оптимізації використання енергії та розробки стратегій управління енергетичними ресурсами в майбутньому. Дослідження підкреслює важливість застосування машинного навчання для інтерпретації тенденцій споживання електроенергії в домогосподарствах, що допомагає ухвалювати обґрунтовані рішення і сприяє сталому використанню енергії. Зокрема, моніторинг споживання електроенергії з використанням алгоритму XG Boost дозволить за допомогою передових аналітичних методів максимізувати використання енергії, підвищити ефективність і надавати споживачам індивідуальні рекомендації для сталих звичок у повсякденному житті.

Стаття [5] розглядає важливість виявлення крадіжок електроенергії (ETD) для забезпечення ефективності розподілу енергії в розумних мережах, підкреслюючи труднощі існуючих методів, які не завжди справляються з великими наборами даних через відсутні значення, варіативність даних і нелінійні взаємозв'язки. Для вирішення цих проблем було розроблено просту, але ефективну модель ETD, яка включає три модулі. Перший модуль застосовує комбінацію методів заповнення пропущених значень, обробки викидів, нормалізації та балансування класів для покращення характеристик часових рядів і створення кращих даних для навчання алгоритмів. Три різні методи машинного навчання, що є незалежними і мають різні підходи до розв'язання цієї проблеми, використовуються як основні моделі навчання. Для покращення точності класифікації на етапі об'єднання результатів алгоритмів машинного навчання застосовується новітній підхід глибинного навчання –

тимчасова конволюційна мережа (TCN). Експериментальні результати показують, що запропонована модель демонструє високу точність і надійність порівняно з іншими добре відомими моделями машинного та глибинного навчання, що робить її практичним інструментом для виявлення крадіжок електроенергії в промислових застосуваннях.

Стаття [6] використовує методи машинного навчання для моделювання місячних даних про споживання електроенергії в Дубаї, Об'єднані Арабські Емірати, на рівні споживача з метою прогнозування споживання на наступний місяць. Основну увагу приділено статистичним багатозмінним підходам, які поєднуються з погодними даними для моделювання споживання електроенергії. Однак у цьому дослідженні також аналізуються дані профілю споживачів разом з регулярними даними про споживання для моделювання та виявлення трендів у споживанні електроенергії. Спочатку два набори даних об'єднуються для створення набору даних, який враховує споживання електроенергії залежно від профілю споживача. Потім застосовується метод відбору ознак на основі кореляції, після чого проводиться процес видалення викидів для отримання очищених даних. Серед різних методів машинного навчання метод Random Forest продемонстрував найкращі результати з найменшими значеннями RMSE та MAPE в процесі 5-кратної кросвалідації.

Стаття [7] присвячена проблемі крадіжок електроенергії, яка є серйозним викликом для енергетичного сектора, зокрема в країнах, як-от Бангладеш, де нелегальне споживання енергії призводить до значних фінансових втрат для постачальників і збільшення витрат на енергетичні послуги для споживачів. Незважаючи на зростання виробництва енергії, крадіжки електрики ставлять під загрозу стабільність енергопостачання та надійність інфраструктури, що негативно впливає на користувачів, які сплачують за споживання енергії законно. Враховуючи важливість запобігання незаконному використанню енергії, це дослідження пропонує методіку, яка застосовує методи машинного навчання для виявлення

нелегальних втрат енергії в незбалансованих наборах даних. Для вирішення проблеми незбалансованих даних використовуються різні методи, зокрема методи надвибірки, як-от випадкова надвибірка, синтетичний метод міноритарних вибірок (SMOTE), підтримувальна векторна машина SMOTE (SVM SMOTE) та граничний SMOTE. Також застосовуються методи зменшення вибірки, такі як випадкове зменшення, центроїд кластеру, редагування найближчого сусіда та правило очищення сусідства. Для аналізу даних споживання, отриманих із вибірки з 56 618 записів за період з 2017 по 2020 роки, використовуються різні методи машинного навчання, зокрема К-найближчий сусід, логістична регресія, метод підтримувальних векторів (SVM) та дерева рішень.

Стаття [8] зосереджена на використанні даних, що генеруються розумними лічильниками, які створюють потоки даних про споживання електроенергії на визначених інтервалах часу. Аналіз цих даних допомагає виявити споживачів з подібними патернами споживання та прогнозувати майбутнє споживання електроенергії, що є важливим для планування виробництва енергії. Однак, оскільки розумні лічильники генерують великі обсяги даних із високою швидкістю, ці дані потребують належного аналізу для отримання корисної інформації. Для цього можна застосовувати такі техніки аналізу, як кластеризація та регресія, які дозволяють виявляти значущі патерни в даних і прогнозувати майбутнє споживання. Існуючі дослідження використовували методи k-середніх та багатолінійної регресії для аналізу даних розумних лічильників за допомогою фреймворку Hadoop, але ці роботи не враховували потік даних про споживання електроенергії. Тому метою цього проєкту є запропонувати систему великих даних під назвою «Прогнозування споживання електроенергії» (PEC), яка використовує фреймворк Spark для аналізу поточних потоків даних.

Стаття [9] підкреслює важливість статистичних методів і штучного інтелекту в сучасному світі, де їх використання стає необхідним, а не лише

теоретичним прикладом. Для задоволення цієї потреби процес оптимізації починається з збору та очищення даних. Метою цієї роботи є надання короткого огляду методів виявлення викидів та пояснення необхідності очищення даних у галузі споживання енергії на прикладі аналізу енергетичного профілю плавального комплексу Технічного університету в Клуж-Напоці. У першій та другій частинах статті надається огляд методів очищення даних, у третій частині порівнюється ефективність запропонованих методів. Наприкінці, в четвертій частині, підведено висновки та обговорюються напрямки для майбутніх досліджень.

У статті [10] аналізуються дані про споживання електроенергії будівлею за період з вересня 2021 по березень 2023 року. Набір даних містить інформацію про споживання електроенергії на 96 часових точках кожного дня, охоплюючи 31 місяць. Метою дослідження є проведення комплексного аналізу поведінки споживання електроенергії будівлею, вивчення тимчасових патернів та виділення потенційних моделей споживання. Спочатку дані піддаються попередній обробці, яка включає очищення даних, обробку відсутніх значень і викидів, а також нормалізацію даних для забезпечення їхньої точності та узгодженості. Далі через візуалізацію даних і статистичний аналіз виявляються періодичність споживання електроенергії в будівлі та характеристики споживання в години пікових навантажень і мінімуму. На основі цього застосовується кластеризація для групування подібних моделей використання енергії та визначення різних характеристик споживання. Результати дослідження надають важливі вказівки для управління енергоспоживанням будівель і оптимізації енергозбереження, що має на меті допомогти операторам будівель ефективніше використовувати енергетичні ресурси, знижувати витрати на споживання енергії та зменшувати вплив на навколишнє середовище. Це дослідження має позитивне значення для сприяння сталому розвитку та розвитку розумних будівель.

Стаття [11] досліджує можливості розвитку смарт-домів, що стають реальністю завдяки технологіям Інтернету речей та машинного навчання. Вони сприяють економії енергії та покращенню зручності життя людей. Для аналізу та прогнозування споживання енергії в смарт-домах використовувалися дані з Kaggle, які містять різні параметри середовища в домогосподарствах. За допомогою моделі машинного навчання було спрогнозовано споживання енергії користувачами, що дозволяє надавати рекомендації щодо заощадження електроенергії. У дослідженні запропоновано схему прогнозування споживання електроенергії в домогосподарствах, яка може бути застосована до прогнозування енергоспоживання в смарт-домах, що підключені до Wi-Fi мережі, а також до регіональних та міських електричних мереж. Це досягнення відкриває нові можливості для використання смарт-домів у майбутньому.

Стаття [12] аналізує та прогнозує споживання електроенергії, що має велике значення для планування енергетичних мереж і управління ресурсами підприємств. Використовуючи історичні дані про споживання електроенергії, зібрані на хмарній платформі системи енергетичного управління Fujian Huatuo, досліджено споживання електроенергії на трьох поверхах компанії Fujian Huatuo Automation Technology. Результати показали, що основними факторами, що впливають на споживання, є погодні умови та робочі дні. Особливу увагу приділено тому, що споживання електроенергії значно змінюється, коли температура перевищує певний поріг. У відповідь на це, у статті запропоновано механізм прогнозування споживання електроенергії на основі кластеризації, що дозволяє передбачати споживання за різними моделями. Експериментальні результати показують, що запропонований механізм ефективно покращує точність прогнозування споживання електроенергії.

Стаття [13] зосереджена на важливості прогнозування споживання енергії в контексті досягнення цілей сталого розвитку, зокрема мети 7, яка

полягає в забезпеченні доступу до надійної, сучасної та доступної енергії для всіх. Швидке зростання міст з високою щільністю населення, а також розвиток комерційного та промислового секторів призводять до збільшення споживання енергії, зокрема у вигляді електроенергії. Прогнозування енергоспоживання є важливим для покращення стратегій управління енергетичними ресурсами в будівлях, зокрема для зменшення енергетичного сліду. У статті проводиться детальний огляд існуючих методів прогнозування енергоспоживання в смарт-будівлях, зокрема для житлових та нежитлових приміщень. Дослідження також розглядає різні категорії навантажень, точність прогнозів та критерії оцінки ефективності методів. Важливими проблемами, що обговорюються, є коливання навантаження, відсутність даних та поведінка споживання електроенергії користувачами. Для зменшення негативного впливу високого енергоспоживання важливо впроваджувати енергоефективні технології та практики в будівлях та промисловості. Хоча впровадження розумних лічильників створює можливості для прогнозування енергоспоживання за допомогою сенсорів і моніторингу енергії, існуючі підходи машинного навчання поки що не враховують хронологічну складність даних. В результаті для прогнозування енергоспоживання запропоновано використання таких методів, як рекурентні нейронні мережі, алгоритми глибинного навчання типу *sequence-to-sequence*, довготривала короткочасна пам'ять (LSTM), згорткові нейронні мережі та гібридні підходи машинного навчання. Ці методи допомагають обробляти відсутні дані за допомогою імпутації середнього значення або використання останніх відомих спостережень, що дозволяє уникнути спотворених оцінок.

Стаття [14] розглядає проблему прогнозування споживання електроенергії в умовах пандемії COVID-19, коли багато урядів ввели обмежувальні заходи, що призвели до значних змін у попиті на електричну енергію. Точне прогнозування споживання електроенергії стає важливим для операторів будівель, щоб покращити ефективність управлінського

планування. Дослідження зосереджується на використанні відкритих онлайн-даних, таких як Google Trends, Google Mobility та COVID-19 дані, для створення точних моделей прогнозування споживання електроенергії в будівлях, що працюють на мікромережах, під час пандемії. Для цього були розроблені моделі екстремального градієнтного бустінгу (XGBoost), регресії на основі підтримувальних векторів (SVR) та авторегресивна інтегрована модель з пояснювальними змінними (ARIMAX). Як кейс-дослідження було проаналізовано реальні дані споживання електроенергії шести поверхової будівлі з мікромережею на території технічного університету в місті Бандунг, Індонезія. Результати показали, що використання онлайн-даних позитивно впливає на точність прогнозування, причому точність зростає, коли використовуються значення з попередніх днів для прогнозування. Моделі XGBoost, які використовують відкладені значення споживання електроенергії, Google Trends та дані про COVID-19 за попередні дні, показали найкращі результати. Проте додавання більшої кількості відкладених предикторів не завжди покращує точність моделей SVR. Моделі ARIMAX, натомість, показали найгірші результати порівняно з іншими моделями.

1.6 Висновки до першого розділу

У розділі були висвітлені ключові аспекти аналізу енергетичних даних, зокрема методи прогнозування споживання електроенергії, які є необхідними для ефективного управління енергетичними системами. ґрунтуючись на змісті цього розділу, можна сформулювати такі висновки:

1) широкий спектр методів для аналізу енергетичних даних. У розділі було охарактеризовано різноманітні методи, що використовуються для аналізу даних у енергетичній сфері, починаючи від класичних статистичних підходів і закінчуючи більш сучасними методами машинного навчання. Це

підтверджує важливість мультидисциплінарного підходу для вирішення задач прогнозування, оптимізації та управління енергетичними потоками;

2) підвищена точність прогнозування за допомогою статистичних і машинних методів. Опис класичних підходів до прогнозування споживання електроенергії, таких як кореляційний та регресійний аналіз, а також використання часових рядів і експоненціального згладжування, демонструє важливість статистичних методів для точного прогнозування, яке необхідне для ефективного планування і управління енергетичними системами. Водночас методи машинного навчання, такі як нейронні мережі та методи глибокого навчання, дозволяють значно покращити точність прогнозів, враховуючи складні взаємозв'язки та нелінійні залежності, що виникають у реальних умовах;

3) важливість адаптації до нових технологій. Розвиток відновлюваних джерел енергії, таких як сонячні та вітрові станції, вимагає інтеграції нових підходів у прогнозуванні та управлінні енергетичними потоками. Застосування методів машинного навчання в прогнозуванні виробництва енергії від цих джерел є критичним для забезпечення стабільності енергетичних мереж і зниження залежності від традиційних джерел енергії;

4) аналіз складних взаємозв'язків в енергетичних системах. Моделі, побудовані на основі машинного навчання, дозволяють не тільки прогнозувати майбутнє споживання енергії, а й аналізувати різні фактори, що впливають на ці процеси, включаючи погодні умови, соціально-економічні зміни та технологічні інновації. Це дає можливість для більш ефективного і гнучкого управління енергетичними системами;

5) перспективи для майбутніх досліджень. Оскільки енергетичні системи стають дедалі складнішими, майбутні дослідження мають зосереджуватися на розвитку більш точних моделей прогнозування і оптимізації, які враховують не лише традиційні фактори, але й більш складні

динамічні зміни, що виникають у результаті інтеграції відновлюваних джерел енергії та змін в енергетичній інфраструктурі;

б) інтеграція різних методів для досягнення більш високої точності. Комбінація статистичних методів, моделей машинного навчання і методів глибокого навчання створює можливості для побудови більш складних і точних систем прогнозування споживання електроенергії, які здатні адаптуватися до змін в умовах енергетичних ринків і потребах споживачів.

Отже, вивчені підходи до аналізу енергетичних даних та методи прогнозування створюють надійну основу для розробки ефективних рішень в енергетичній галузі, які дозволяють знижувати витрати на енергію, оптимізувати використання енергетичних ресурсів та забезпечувати стабільну роботу енергетичних систем в умовах змінного попиту.

РОЗДІЛ 2. ДОСЛІДЖЕННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ СТВОРЕННЯ МОДЕЛІ АНАЛІЗУ ДАНИХ СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ

2.1 Техніки навчання в енергетичних даних

Техніки машинного навчання (рис. 2.1) стали незамінним інструментом у багатьох галузях, зокрема в енергетичному секторі, де ці методи використовуються для прогнозування, оптимізації та аналізу великих обсягів даних. Зокрема, два основних підходи до машинного навчання – навчання з учителем та навчання без учителя – є важливими для вирішення специфічних задач, які виникають у контексті енергетичних даних [6]. Кожен із цих підходів має свої переваги та обмеження, що визначає їхнє застосування в залежності від специфіки задачі, доступних даних та вимог до точності та швидкості результатів.

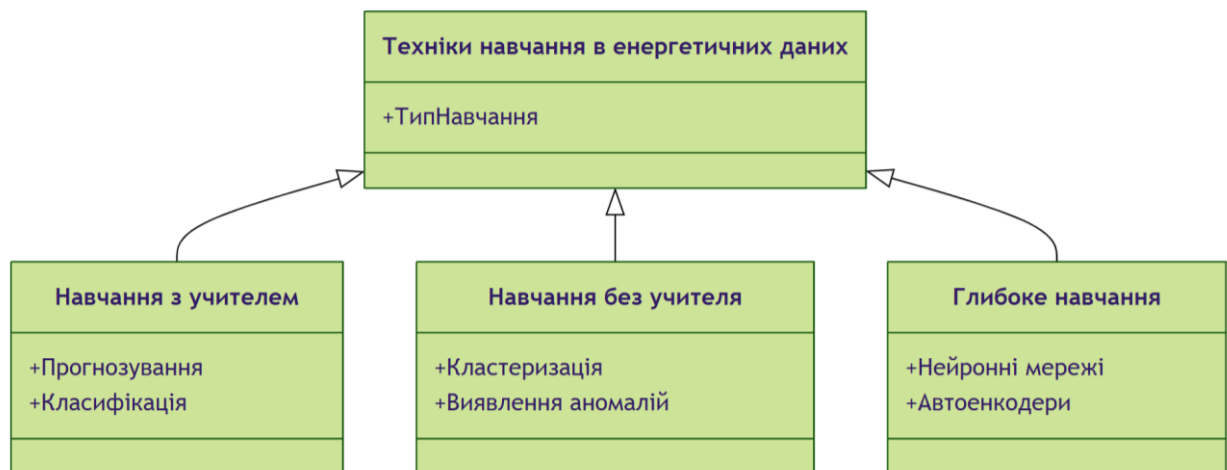


Рисунок 2.1 – Техніки навчання в енергетичних даних

Навчання з учителем (supervised learning) є одним із найбільш широко використовуваних методів машинного навчання, при якому система

навчається на основі наданих їй вхідних даних, що супроводжуються відповідними мітками або правильними відповідями. В енергетичному секторі цей підхід застосовується для вирішення задач, де є доступ до історичних даних про споживання енергії, а також інших факторів, що можуть впливати на процеси енергетичного виробництва чи споживання. Ключовим аспектом є наявність вхідних та вихідних даних, що дозволяє здійснювати навчання моделі, орієнтуючи її на конкретну задачу. Одним з основних застосувань навчання з учителем в енергетичному секторі є прогнозування споживання електричної енергії. В цьому випадку вхідними даними можуть бути різні фактори, такі як температура повітря, час доби, сезонні коливання, історичні дані про споживання енергії, економічні показники або навіть соціальні аспекти (наприклад, святкові дні чи вихідні). Вихідними даними є значення споживаної енергії за відповідний період часу. На основі цих даних будуються моделі, що дозволяють прогнозувати майбутнє споживання, роблячи можливим планування енергетичних мереж і оптимізацію їхніх ресурсів.

Методи, що застосовуються в навчанні з учителем, включають регресію та класифікацію. Регресія використовується для прогнозування числових значень, таких як кількість електроенергії, яку буде спожито за певний період, в той час як класифікація може бути використана для розподілу даних на категорії, наприклад, для класифікації днів за рівнем споживання енергії (пік, середнє, низьке). Одним із популярних алгоритмів навчання з учителем є метод опорних векторів (SVM), який застосовується для пошуку оптимальної гіперплощини, що розділяє дані на класи або мінімізує помилки в прогнозах. Серед інших методів, які активно використовуються в енергетичних задачах, можна виділити дерева рішень, випадкові ліси та градієнтний бустинг. Такі методи дозволяють враховувати складні взаємозв'язки між різними змінними та забезпечують високу точність прогнозів.

Також важливим напрямком є застосування нейронних мереж, зокрема глибоких нейронних мереж, які здатні працювати з великими обсягами даних

і складними, нелінійними залежностями. Завдяки своїй здатності виявляти приховані закономірності в даних, нейронні мережі дозволяють створювати високоточні моделі для прогнозування споживання енергії в реальному часі, враховуючи численні фактори, такі як погодні умови, економічні коливання або зміни в поведінці споживачів.

Прогнозування виробництва енергії від відновлюваних джерел, таких як вітрові або сонячні станції, є ще однією важливою задачею, де активно використовуються методи навчання з учителем. Вітрові турбіни, наприклад, виробляють електричну енергію залежно від швидкості вітру, а сонячні панелі – від рівня сонячної радіації. Знання про ці параметри дозволяє з високою точністю прогнозувати енергетичне виробництво, що є важливим для забезпечення стабільності мережі та оптимізації використання ресурсів.

Навчання без учителя (*unsupervised learning*) є іншим підходом, який застосовується в машинному навчанні, коли дані не містять явних міток або відповідей, і задача полягає в тому, щоб виявити структури, закономірності чи класи в даних без попереднього визначення результату [6, 7]. У енергетичній галузі цей підхід використовується для виявлення аномалій, кластеризації споживачів, а також для сегментації великих масивів даних, що дозволяє здійснювати глибший аналіз.

Одним з найпоширеніших методів навчання без учителя є кластеризація. У контексті енергетичних даних цей метод дозволяє групувати споживачів електричної енергії на основі схожих характеристик, таких як регулярність споживання, час пікового навантаження або кількість енергії, що споживається. Наприклад, можна виділити класи споживачів, які мають подібні патерни споживання, що дозволяє більш точно налаштувати енергетичні мережі та прогнози для різних груп споживачів.

Також методи без учителя активно застосовуються для виявлення аномалій у даних. Аномальні ситуації в енергетичних мережах, такі як несподіване підвищення або зниження споживання енергії, можуть бути

ознакою несправностей, атак або порушень у роботі обладнання. Алгоритми, такі як алгоритм К-середніх або методи на основі випадкових лісів, здатні виявляти ці аномалії, що дозволяє своєчасно вжити заходів для виправлення ситуації, забезпечивши стабільність роботи енергетичних систем. Для більш глибокого аналізу та кластеризації в енергетичних даних застосовуються також методи зниження розмірності, такі як головні компоненти (РСА) або методи, що базуються на автоенкодерах у нейронних мережах. Ці методи дозволяють зменшити кількість змінних, зберігаючи при цьому основні закономірності в даних, що значно полегшує їх подальший аналіз та використання.

Навчання без учителя також може бути використане для аналізу даних про відновлювальні джерела енергії, де є потреба виявити непередбачувані закономірності або аномалії в змінних умовах, таких як зміни в швидкості вітру або рівні сонячної радіації. Методи без учителя дозволяють більш гнучко адаптуватися до нових умов, що є важливим в умовах постійно змінюваних параметрів навколишнього середовища.

Основною відмінністю між навчанням з учителем та без учителя є спосіб взаємодії з даними. У навчанні з вчителем для кожного елемента даних є відповідна мітка або цільова змінна, тоді як у навчанні без учителя система сама повинна знаходити структури в даних [8]. Обидва підходи можуть бути використані в енергетичних задачах, але вибір залежить від наявності міток, складності задачі та типу даних.

Таким чином, навчання з вчителем часто застосовується в ситуаціях, коли потрібно передбачити конкретний результат (наприклад, споживання енергії) на основі визначених вхідних змінних, а навчання без учителя є більш ефективним для виявлення прихованих закономірностей та класифікації великих обсягів незамічених даних, таких як кластеризація споживачів або виявлення аномалій. Техніки навчання з учителем та без учителя є важливими інструментами для аналізу енергетичних даних. Вибір конкретного підходу

залежить від типу задачі, доступності міток та бажаної точності результатів. Застосування цих методів дозволяє значно покращити ефективність енергетичних систем, знизити витрати на енергетичні ресурси, а також підвищити надійність і стабільність роботи енергетичних мереж, забезпечуючи оптимальне використання наявних ресурсів.

2.2 Використання нейронних мереж у прогнозуванні споживання електроенергії

Використання нейронних мереж у прогнозуванні споживання електроенергії є одним із найбільш перспективних та ефективних напрямків у галузі машинного навчання та штучного інтелекту. Сучасні нейронні мережі, завдяки своїй здатності до самонавчання і адаптації до складних, нелінійних закономірностей, дозволяють вирішувати складні завдання, що виникають у процесі прогнозування енергетичного попиту. Прогнозування споживання електроенергії є надзвичайно важливим завданням для енергетичних компаній, оскільки воно дозволяє оптимізувати роботу енергетичних мереж, знижувати витрати на генерацію електричної енергії та зменшувати ризики від енергетичних дефіцитів або надлишків.

Нейронні мережі використовуються для прогнозування споживання електроенергії, оскільки вони здатні ефективно моделювати складні залежності, які виникають між різноманітними змінними, що впливають на рівень споживання [8, 9]. До таких змінних можна віднести погодні умови, сезонні коливання, економічні фактори, соціальні аспекти, а також особливості самих споживачів. Для того щоб зрозуміти, чому нейронні мережі є настільки ефективними у цьому контексті, потрібно розглянути основи їхнього функціонування, різні типи нейронних мереж та їхнє застосування у специфічних задачах прогнозування в енергетичному секторі. Нейронні мережі за своєю природою є математичними моделями, натхненими

біологічними нейронами. Вони складаються з численних вузлів, або нейронів, які зв'язуються між собою через зважені синапси. Кожен нейрон отримує на вхід певну кількість сигналів (даних), обробляє їх за допомогою певної функції активації і передає результат наступному нейрону. Така архітектура дозволяє нейронним мережам адаптуватися до складних залежностей і виявляти закономірності, які важко помітити за допомогою традиційних математичних методів. Одним із найбільш популярних типів нейронних мереж, які використовуються для прогнозування споживання електроенергії, є багатошарові перцептрони (MLP). Ці мережі складаються з декількох шарів нейронів, де кожен наступний шар отримує сигнали від попереднього і передає їх на вихід. MLP використовуються для вирішення задач регресії та класифікації, що робить їх ефективними для прогнозування числових значень споживання енергії. Основним методом навчання багатошарових перцептронів є метод зворотного поширення помилки (backpropagation), який дозволяє коригувати ваги між нейронами на основі помилок, що виникають при порівнянні прогнозованих і фактичних значень. Іншим важливим типом нейронних мереж, який активно застосовується у прогнозуванні споживання електроенергії, є рекурентні нейронні мережі (RNN). Ці мережі мають особливість, що дозволяє їм працювати з послідовними даними, зберігаючи інформацію про попередні етапи [10]. Це робить RNN особливо корисними для задач прогнозування, де важливу роль відіграють часові залежності, як це часто буває у задачах прогнозування попиту на електричну енергію. Прогнозування споживання електроенергії зазвичай є задачею, яка включає часові ряди, оскільки споживання змінюється залежно від часу доби, сезону, економічної ситуації та багатьох інших факторів. Рекурентні нейронні мережі дозволяють враховувати такі тимчасові залежності, що робить їх дуже потужними для прогнозування енергетичних потреб.

Одним з найбільш передових підходів у використанні нейронних мереж для прогнозування енергетичних даних є застосування довготривалих

короткочасних пам'ятей (LSTM). LSTM є модифікацією рекурентних нейронних мереж, яка дозволяє ефективно працювати з довгими послідовностями даних, не забуваючи важливу інформацію з попередніх етапів. Ці мережі мають особливість підтримувати «довготривалу пам'ять», що дозволяє їм краще адаптуватися до задач, де потрібно враховувати віддалені часові залежності, наприклад, сезонні коливання споживання електроенергії або довгострокові тенденції в зміні попиту. Одним із основних напрямків застосування нейронних мереж у прогнозуванні споживання електроенергії є прогнозування попиту в реальному часі. Це завдання особливо актуальне для енергетичних компаній, які повинні оперативного регулювати виробництво та споживання енергії. Використання нейронних мереж дозволяє не лише здійснювати прогнозування на основі історичних даних про споживання, а й враховувати фактори, які можуть змінюватися в режимі реального часу [10, 11]. Наприклад, зміни в погодних умовах можуть значно вплинути на споживання електроенергії, оскільки в зимовий період збільшується споживання тепла, а влітку – охолодження.

Нейронні мережі також використовуються для прогнозування попиту на електроенергію на більш довгостроковій основі. Наприклад, нейронні мережі можуть допомогти прогнозувати споживання в умовах зміни економічної ситуації, зростання або зменшення населення, розвитку нових технологій та інфраструктури. Таке прогнозування є важливим для планування енергетичної інфраструктури та розробки стратегій щодо сталого розвитку енергетичних систем. Крім того, нейронні мережі активно використовуються для оптимізації роботи енергетичних мереж. Наприклад, вони можуть допомогти в оптимізації розподілу електричної енергії між різними районами, забезпечуючи більш ефективне використання енергетичних ресурсів і знижуючи витрати. Нейронні мережі можуть також застосовуватися для виявлення аномалій у роботі енергетичних мереж, таких як несправності в обладнанні або несанкціоноване споживання енергії. Це дозволяє своєчасно виявляти

проблеми та вживати необхідні заходи для їхнього усунення [11, 12]. Одним із основних переваг нейронних мереж є їх здатність до самооптимізації та адаптації до змінних умов. На відміну від традиційних статистичних методів, нейронні мережі можуть враховувати велику кількість параметрів та факторів, не вимагаючи явного програмування для кожного з них. Вони також здатні виявляти складні нелінійні залежності між різними змінними, що робить їх надзвичайно потужними для вирішення задач прогнозування споживання електроенергії.

Потрібно також зазначити, що успіх використання нейронних мереж для прогнозування споживання електроенергії значною мірою залежить від якості даних, які використовуються для навчання. Оскільки нейронні мережі є методами навчання з великою кількістю параметрів, вони потребують великих обсягів історичних даних для досягнення високої точності прогнозів. Це включає дані про споживання електроенергії, погодні умови, соціально-економічні фактори та інші параметри. Чим більше якісних даних доступно для навчання, тим точнішими і надійнішими будуть результати прогнозування.

Таким чином, нейронні мережі є надзвичайно потужним інструментом для прогнозування споживання електроенергії. Їх здатність обробляти великі обсяги даних, враховувати складні залежності та адаптуватися до змінних умов робить їх незамінними в енергетичному секторі. Використання таких методів дозволяє значно покращити точність прогнозів, оптимізувати використання енергетичних ресурсів, підвищити ефективність енергетичних мереж і забезпечити стабільне постачання енергії споживачам.

2.3 Методи очищення та підготовки даних для аналізу споживання електроенергії

Методи очищення та підготовки даних для аналізу споживання електроенергії є невід'ємною частиною кожного процесу аналізу великих масивів даних, зокрема в енергетичній галузі, де якість даних безпосередньо впливає на ефективність прогнозування, оптимізації та управління енергетичними потоками [13]. Успіх моделювання, прогнозування та прийняття рішень на основі даних значною мірою залежить від того, наскільки коректно та систематизовано підготовлені вхідні дані. Процес очищення і підготовки даних передбачає цілу низку дій, які спрямовані на усунення різноманітних видів шуму, помилок та невідповідностей у даних, а також на приведення їх до зручного для аналізу виду. Зокрема, в контексті аналізу споживання електроенергії ці методи мають ще більшу актуальність, оскільки вони безпосередньо впливають на точність прогнозів, що можуть бути використані для оптимізації енергетичних мереж, моніторингу споживання та управління енергетичними ресурсами. Одним із найбільш критичних аспектів у підготовці даних для аналізу споживання електроенергії є вирішення проблеми неповних або відсутніх даних. Відсутність даних, чи то через технічні помилки, збій у вимірювальних пристроях або неправильне зчитування показників, є загальноприйнятною проблемою, з якою стикаються аналітики. У таких випадках необхідно застосовувати методи імпутації, які дозволяють заповнити прогалини в даних на основі інших відомих значень. Це може бути зроблено за допомогою середніх значень, медіани, або більш складних методів, таких як регресія чи інтерполяція, де пропущені значення обчислюються на основі співвідношення між сусідніми спостереженнями. У разі прогнозування споживання електроенергії важливо забезпечити, щоб дані, використані для моделювання, були повними і точними, оскільки навіть незначні помилки можуть призвести до значних відхилень у результатах. Крім

того, часто виникає проблема некоректних або аномальних значень, які можуть бути викликані технічними збоєм або іншими факторами, що порушують нормальний хід вимірювань. Такі значення можуть суттєво спотворювати результати аналізу, тому необхідно застосовувати методи виявлення та корекції аномалій [12-14]. Одна з найбільш популярних стратегій полягає у використанні методів виявлення викидів, які полягають у порівнянні кожного значення з середнім або медіанним значенням за певний період. Якщо значення сильно відрізняється від нормального діапазону, воно може бути класифіковане як аномальне, і в такому випадку його слід замінити на середнє значення або інтерпольовані дані. Іншою поширеною технікою є використання методів машинного навчання для виявлення аномалій, де спеціалізовані моделі навчаються розпізнавати відхилення, що виходять за межі стандартного функціонування системи.

Також важливою частиною процесу підготовки даних є нормалізація та стандартизація. Оскільки дані, зібрані з різних джерел, можуть мати різні масштаби, необхідно привести їх до єдиного виду. Це особливо актуально в задачах машинного навчання, де моделі можуть бути чутливими до розмірів значень вхідних змінних. Нормалізація передбачає масштабування даних до певного діапазону, зазвичай від 0 до 1, або стандартизацію, коли дані приводяться до нульового середнього значення з одиничною дисперсією. Очищення даних від шуму і приведення їх до єдиного масштабу дозволяє значно підвищити точність моделей, що використовуються для прогнозування.

Після того, як дані очищено від аномалій та нормалізовано, наступним важливим кроком є їхня трансформація. Враховуючи, що дані про споживання електроенергії часто мають часовий характер, використання часових рядів є невід'ємною частиною аналізу. Дані зберігають певну структуру, де значення в різний час можуть бути залежними одне від одного, що означає необхідність у додаткових перетвореннях, таких як створення змінних лагу, сезонних

компонентів або трендів [13, 14]. Трансформація таких даних дозволяє моделювати залежності, які існують між різними часовими точками, що є важливим для ефективного прогнозування споживання в майбутньому. Одним з важливих аспектів підготовки даних для прогнозування є також категоризація та агрегування даних. Наприклад, для аналізу споживання електроенергії за певний період можна агрегувати дані по годинах, днях, тижнях або місяцях, залежно від того, яка інформація є більш важливою для конкретної задачі. Таке агрегування дозволяє зменшити обсяг даних, зберігаючи при цьому важливі тенденції та властивості. У багатьох випадках, для спрощення аналізу, використовуються категоріальні змінні, які дозволяють класифікувати споживання за певними критеріями, наприклад, за типом споживача, регіоном чи погодними умовами. Це дозволяє зосередитися на ключових аспектах даних і покращити точність прогнозів. Особливо важливою є перевірка якості даних після очищення та підготовки [14]. Це передбачає використання різноманітних методів візуалізації та статистичних тестів, які дозволяють оцінити, наскільки добре дані відображають реальний процес споживання електроенергії, а також виявити будь-які інші проблеми, які можуть залишитися після попередніх етапів. Візуалізація може включати графіки, що показують тенденції споживання електроенергії, а також гістограми та коробкові діаграми для перевірки розподілу даних і виявлення потенційних аномалій. Важливо також враховувати взаємозв'язок між різними параметрами споживання електроенергії. Наприклад, погодні умови можуть мати істотний вплив на попит на електроенергію, оскільки в холодну погоду споживання енергії зростає через необхідність опалення, а влітку – через кондиціонування повітря. Тому при підготовці даних для аналізу важливо враховувати ці фактори, а також зв'язок між ними. Для цього застосовуються методи створення нових змінних, які дозволяють врахувати взаємозалежність різних факторів.

Таким чином, очищення та підготовка даних є складним і багатоступеневим процесом, що включає кілька етапів, на кожному з яких виконуються специфічні операції для підвищення якості даних. Це дозволяє створювати надійні моделі для прогнозування споживання електроенергії, що є необхідним для ефективного управління енергетичними ресурсами, оптимізації виробництва і розподілу енергії, а також для забезпечення стабільності та ефективності енергетичних мереж.

2.4 Проблеми та особливості роботи з великими даними в енергетичній сфері

Проблеми та особливості роботи з великими даними в енергетичній сфері є актуальними в умовах сучасного розвитку технологій, які дають змогу збирати та аналізувати величезні обсяги даних [13]. З розвитком інтелектуальних електричних мереж (smart grids), впровадженням автоматизованих систем управління та збору даних, а також використанням високочастотних вимірювань для моніторингу споживання електроенергії, з'являються нові можливості для підвищення ефективності енергетичних систем. Однак з великою кількістю даних постають значні технічні та аналітичні труднощі, що вимагають розв'язання специфічних проблем, пов'язаних з обробкою, зберіганням, аналізом та інтерпретацією таких великих обсягів інформації. Однією з основних проблем роботи з великими даними в енергетичній сфері є обсяг та різноманіття даних. Системи автоматизованого збору даних можуть генерувати потоки інформації в реальному часі, що включають вимірювання електричного споживання, стан обладнання, параметри навколишнього середовища (наприклад, температура, вологість, швидкість вітру), дані про економічну ситуацію, тарифи, а також інформацію про соціальні чинники, які можуть впливати на рівень споживання енергії. Усі ці дані можуть бути як структурованими, так і неструктурованими,

і їх величезний обсяг ставить перед фахівцями з аналізу даних завдання зберігання та обробки такої інформації. Це також призводить до необхідності використання потужних обчислювальних ресурсів, здатних обробляти ці дані без втрат ефективності [14]. Ще однією проблемою є різноманіття джерел та типів даних. В енергетичній сфері дані можуть бути зібрані з різних датчиків, сенсорів, рахунків, а також з інших джерел, таких як зовнішні метеорологічні служби або соціально-економічні бази даних. Це призводить до проблеми інтеграції даних з різних систем і платформ, оскільки ці дані можуть мати різні формати, одиниці вимірювання або часові мітки. Для коректного аналізу даних важливо забезпечити їхнє уніфіковане представлення, що включає нормалізацію, очищення та приведення до єдиного формату. Проблеми з інтеграцією та консолідацією даних можуть ускладнити процес їхнього подальшого використання для прогнозування або управління енергетичними процесами.

Крім того, аналіз великих даних у енергетичній сфері часто стикається з проблемою якості даних. Низька якість даних є значною проблемою, яка може виникнути через некоректне вимірювання, помилки в роботі датчиків або збої в комунікаційних каналах. Погана якість даних може також виникати через пропущені або недостовірні значення, які необхідно ідентифікувати та коригувати до початку аналізу. Для цього використовуються різноманітні методи очищення даних, включаючи виявлення та заміну аномальних або пропущених значень, а також фільтрацію шуму, що може спотворювати результати аналізу. Процес очищення даних є критично важливим для забезпечення точності прогнозів та моделей, що використовуються для оптимізації роботи енергетичних систем.

Окремим викликом є проблема зберігання великих обсягів даних. Залежно від типу та обсягу інформації, енергетичні компанії можуть стикатися з необхідністю зберігати дані у великих масштабах, що вимагає використання ефективних засобів для зберігання та обробки даних, таких як хмарні

обчислення або спеціалізовані бази даних для великих даних (big data). Важливо, щоб такі системи зберігання були здатні ефективно управляти великими обсягами даних, забезпечуючи високу швидкість доступу до інформації та можливість її обробки в реальному часі [7, 11]. Для цього розробляються спеціалізовані платформи для обробки великих даних, такі як Hadoop чи Apache Spark, які дозволяють розподіляти обробку даних між численними обчислювальними вузлами, що сприяє значному зменшенню часу, необхідного для їх аналізу.

Аналіз великих даних у енергетичній сфері також стикається з проблемами пов'язаними з часом та частотою збору даних. У реальному часі системи повинні бути здатні обробляти інформацію, яка надходить від численних датчиків і приладів, і враховувати мінливі умови, такі як зміни в попиті на електричну енергію або коливання в якості мережі. Це потребує високошвидкісних алгоритмів обробки даних, що дозволяють миттєво реагувати на зміни в системі і здійснювати прогнози для оптимізації роботи енергетичної мережі. Висока частота збору даних, зокрема для моніторингу в режимі реального часу, може стати проблемою для систем з обмеженими обчислювальними потужностями або коли йдеться про обробку потокових даних.

Не менш важливим є аспект безпеки та конфіденційності даних. В умовах цифровізації енергетичних мереж виникає необхідність захисту даних від несанкціонованого доступу, що може призвести не тільки до порушення конфіденційності, а й до можливих збоїв у роботі мереж або навіть до кібератак. Захист великих обсягів даних вимагає застосування високотехнологічних систем безпеки, які включають шифрування, багаторівневу автентифікацію та інші методи, що гарантують збереження конфіденційності інформації. Водночас необхідно забезпечити доступ до важливих даних для аналітиків та інших зацікавлених осіб без порушення вимог безпеки. Ще одним викликом, який виникає під час роботи з великими

даними в енергетичній сфері, є відсутність єдиних стандартів для збирання, зберігання та обробки даних. На сьогоднішній день існує велика кількість різних платформ і технологій, що використовуються для моніторингу та обробки даних в енергетичних мережах, і вони можуть не бути сумісними між собою [13]. Це створює проблему для інтеграції різних систем і платформ, оскільки кожен елемент інфраструктури може використовувати власний формат даних, який не завжди підтримується іншими системами. Для ефективної роботи з великими даними необхідно створювати стандарти, які дозволяють інтегрувати різні джерела інформації та забезпечити безперешкодний обмін даними між системами.

Незважаючи на ці проблеми, робота з великими даними в енергетичній сфері має величезний потенціал. Завдяки використанню новітніх технологій для обробки та аналізу даних можна значно покращити прогнозування попиту на електроенергію, оптимізувати роботу енергетичних мереж, знижувати витрати на виробництво енергії та підвищувати ефективність використання ресурсів. Розробка нових методів обробки та зберігання великих даних відкриває можливості для створення більш адаптивних, надійних і економічно ефективних енергетичних систем, що здатні реагувати на зміни попиту, мінливі погодні умови, соціально-економічні фактори та інші змінні.

Таким чином, успіх роботи з великими даними в енергетичній сфері залежить від здатності ефективно вирішувати проблеми з їх обробкою, зберіганням, інтеграцією та безпекою. Враховуючи високий темп розвитку технологій, постійно зростаючі обсяги зібраних даних та їхню важливість для прийняття рішень у реальному часі, розв'язання цих проблем має стати основним пріоритетом для енергетичних компаній і державних органів. Оскільки обробка великих даних відкриває нові можливості для розвитку інтелектуальних енергетичних мереж, важливим завданням є створення нових технологій та платформ, які дозволяють ефективно працювати з такими

даними, мінімізуючи ризики та максимізуючи можливості для вдосконалення енергетичних систем.

2.5 Аналіз залежності споживання електроенергії від зовнішніх факторів

Аналіз залежності споживання електроенергії від зовнішніх факторів, таких як погодні умови, сезонність та економічні фактори, є важливою складовою досліджень в енергетичній галузі. Споживання електричної енергії є складним процесом, який залежить від численних чинників, що можуть змінюватися в часі, і тому вивчення цих залежностей є важливим для прогнозування попиту на енергію, оптимізації її виробництва та зниження енергетичних витрат [12-14]. Зовнішні фактори, як погодні умови, сезонні коливання або зміни в економічній ситуації, можуть мати суттєвий вплив на рівень споживання електроенергії, що в свою чергу позначається на роботі енергетичних систем, тарифах, ефективності розподілу та передачі енергії. Оскільки рівень споживання електричної енергії є однією з основних характеристик для планування роботи енергетичних підприємств, розуміння того, як ці фактори впливають на попит, має важливе значення для забезпечення стабільності енергетичних систем і економічної ефективності.

Погода є одним із найбільш значущих зовнішніх факторів, який прямо впливає на споживання електроенергії. Температурні коливання, вологість, швидкість вітру, кількість опадів – всі ці показники можуть змінювати рівень споживання електричної енергії, оскільки в холодні місяці потреба в обігріві приміщень зростає, а в літній період – в охолодженні повітря. При температурних коливаннях підвищується споживання електроенергії в будівлях через збільшення використання опалення або кондиціонерів. Наприклад, в умовах морозних зимових місяців попит на електроенергію зростає через збільшення використання електричних обігрівачів, тоді як влітку – через використання кондиціонерів і вентиляторів. Тому погодні умови

суттєво впливають на споживання енергії не лише в побутовому секторі, але й в промисловості, де збереження температурного режиму є важливим фактором для безперебійної роботи виробництв. Більш того, часті погодні явища, такі як бурі або сильні дощі, можуть спричиняти перебої в електропостачанні, що також потрібно враховувати в розрахунках споживання енергії на різних етапах.

Оскільки зміни в погодних умовах є циклічними, сезонність також відіграє важливу роль в аналізі залежності споживання електроенергії. Виходячи з сезонних коливань температури, споживання електричної енергії має яскраво виражену сезонну залежність. У зимові місяці споживання зазвичай зростає через підвищену потребу в опаленні, а влітку, навпаки, зростає потреба в охолодженні приміщень [8, 10]. Проте, окрім температурних коливань, сезонність також впливає на рівень споживання через зміни в умовах освітлення. У зимовий період світловий день значно коротший, тому споживання електроенергії зростає не лише через обігрів, але й через використання електричних приладів для освітлення. У літній період, коли світловий день довший, попит на електричну енергію може бути меншим, оскільки використання освітлення знижується, однак попит на енергію для кондиціонування може компенсувати це зниження.

Сезонні зміни також мають значення для сільськогосподарських, промислових та інших секторів економіки, де виробничі процеси та потреба в енергії можуть змінюватися в залежності від пори року. Наприклад, в аграрному секторі влітку відбуваються пікові навантаження на енергетичні мережі через потребу в іригації та переробці сільськогосподарських культур, а в зимовий період – через зростання попиту на електроенергію для теплиць і зберігання продукції.

Економічні умови є ще одним важливим чинником, що впливає на рівень споживання електричної енергії. В умовах економічного зростання, зазвичай спостерігається збільшення попиту на енергію через підвищення виробничих

обсягів у промисловості, зростання рівня споживання енергії в побутовому секторі через збільшення рівня доходів населення, а також через підвищену активність у різних секторах економіки, таких як транспорт, комунальні послуги та інші [11]. У періоди економічної кризи або рецесії спостерігається зниження попиту на енергію, оскільки зменшується обсяг виробництва в промисловості, спостерігається зниження рівня споживання в побуті через зменшення доходів населення і зниження рівня активності в інших секторах економіки. Це означає, що економічні цикли безпосередньо впливають на рівень споживання енергетичних ресурсів, оскільки зменшення або збільшення економічної активності спричиняє пропорційні зміни в попиті на електричну енергію.

Незважаючи на значний вплив цих зовнішніх факторів на рівень споживання електричної енергії, їхня взаємодія є складною та багатофакторною. Наприклад, в умовах зміни економічних умов, ефекти, що викликаються погодними умовами, можуть бути посилені або пом'якшені. У цей період особливо важливо розуміти, як саме змінюються взаємозв'язки між цими факторами, щоб зробити прогнози більш точними і надійними.

Для того, щоб успішно прогнозувати споживання електроенергії, важливо застосовувати комплексний підхід, який би враховував усі ці фактори в їхній взаємодії. Для цього широко використовуються математичні моделі, які дозволяють будувати залежності між споживанням енергії та погодними умовами, економічними чинниками, а також іншими зовнішніми впливами. Одним із найпоширеніших підходів є використання моделей регресії, нейронних мереж, методів машинного навчання та статистичних методів для аналізу цих залежностей [12]. Такі методи дозволяють не лише оцінювати поточний стан споживання електроенергії, але й будувати прогнози на основі різноманітних факторів, що мають вплив на енергетичну систему.

Таким чином, аналіз залежності споживання електроенергії від зовнішніх факторів дозволяє отримати цінну інформацію для планування

енергетичних мереж, оцінки ефективності роботи електричних систем, а також для розробки стратегій з оптимізації виробництва та споживання енергії. Враховуючи циклічний характер погодних умов і економічних змін, а також їх взаємодію, можна знизити ризики перебоїв у постачанні електричної енергії, забезпечити сталість енергетичних систем і забезпечити енергетичну безпеку в умовах змінного попиту.

2.6 Вибір методу навчання

Вибір методу навчання є ключовим етапом у розв'язанні задач прогнозування та класифікації, де важливо правильно підібрати алгоритм, який би ефективно працював з наявними даними [12-14]. Методи регресії, класифікації та нейронні мережі – це три основні категорії методів машинного навчання, які мають широкий спектр застосування в різних галузях, зокрема в енергетичній сфері для прогнозування споживання електроенергії, виявлення аномалій в даних та оптимізації процесів. Вибір між цими методами залежить від характеру задачі, типу даних, вимог до точності та складності обчислень, що є важливими факторами при прийнятті рішення про метод навчання. Кожен з цих методів має свої переваги та обмеження, що визначають їх ефективність у різних контекстах. У цьому контексті, порівняння цих методів є важливим для розуміння того, коли і чому варто вибрати один із них, а також для забезпечення максимальної ефективності в розв'язанні конкретної задачі.

Регресія є одним з найстаріших і найбільш відомих методів статистичного аналізу, який використовується для прогнозування числових значень на основі вхідних даних. Основною метою регресії є встановлення математичної залежності між вхідними змінними і вихідною змінною, що дозволяє прогнозувати значення останньої. Регресія може бути лінійною або нелінійною, залежно від характеру зв'язку між змінними. Лінійна регресія, яка

є найбільш простою і найпоширенішою, використовується тоді, коли передбачається, що залежність між змінними є лінійною, тобто вихідна змінна змінюється пропорційно до змін вхідних змінних. Нелінійні методи регресії використовуються, коли залежність між змінними є складнішою, і їх застосування може бути більш ефективним для моделювання реальних ситуацій, де лінійна залежність не підходить. Перевагою методу регресії є його простота, інтерпретованість результатів і швидкість навчання, оскільки регресія не потребує значних обчислювальних ресурсів, що робить її корисною для задач, де важлива швидкість розрахунків. Однак одним з обмежень регресії є її неспроможність адекватно справлятися з більш складними та нелінійними залежностями, особливо в тих випадках, коли кількість факторів впливу на результат є великою і важко побудувати явну формулу для зв'язку між змінними [10, 14].

Класифікація, на відміну від регресії, є задачею, в якій вихідна змінна є категоріальною, тобто результатом є одна з множини класів або категорій. Основною метою класифікації є визначення, до якого класу належить конкретний об'єкт або спостереження, виходячи з його характеристик. Класифікація може бути двокласовою або багатокласовою, залежно від кількості можливих результатів. Алгоритми класифікації можуть бути різними: від простих, таких як наївний байєсівський класифікатор або метод опорних векторів, до більш складних, таких як деревовидні структури або багатосарові нейронні мережі. Одним з найбільш розповсюджених методів класифікації є метод опорних векторів (SVM), який здобув популярність завдяки своїй здатності ефективно працювати з високовимірними даними і забезпечувати високу точність в задачах класифікації, що не піддаються простим лінійним моделям. Однією з переваг класифікації є її здатність до точного визначення класів на основі вхідних ознак, навіть коли дані є шумними або неповними. Однак класифікація, як і регресія, має обмеження в контексті складності залежностей між ознаками, що може потребувати

використання більш складних і ресурсозатратних алгоритмів для досягнення високої точності, особливо коли мова йде про багатокласові задачі або великі набори даних.

Нейронні мережі являють собою складні математичні моделі, натхненні біологічними процесами, що використовуються для вирішення задач як регресії, так і класифікації. Нейронні мережі, зокрема багат шарові перцептрони, є здатними до навчання складних функцій, які не можуть бути виявлені за допомогою традиційних методів, таких як регресія або класифікація. Вони відрізняються великою гнучкістю та здатністю до апроксимації складних, нелінійних зв'язків між змінними [14]. Нейронні мережі можуть бути використані для вирішення широкого спектра задач, включаючи як прогнозування числових значень (регресія), так і класифікацію об'єктів. Однією з основних переваг нейронних мереж є їх здатність до ефективного навчання складних нелінійних залежностей між вхідними даними та результатом, що дозволяє їм бути більш точними в задачах, де інші методи, зокрема регресія та класифікація, можуть зазнавати труднощів. Однак вони також мають свої обмеження, серед яких найбільшими є потреба в значних обчислювальних ресурсах, вимога до великих обсягів навчальних даних і складність у налаштуванні параметрів, таких як кількість шарів або нейронів у кожному шарі, що робить процес навчання часозатратним. Крім того, нейронні мережі є менш інтерпретованими, ніж регресія та класифікація, що може створювати труднощі при необхідності аналізу або пояснення отриманих результатів.

Порівняння цих трьох методів показує, що кожен з них має свої сильні та слабкі сторони. Регресія, завдяки своїй простоті та швидкості, є найбільш підходящою для задач з лінійними залежностями, де є достатньо мало змінних і модель повинна бути швидко побудована. Класифікація є більш потужним інструментом для задач, де результат є категоріальним, а також для більш складних і високовимірних даних, однак вона вимагає більш складних

алгоритмів і може мати обмеження в разі великої кількості класів. Нейронні мережі є універсальними і можуть бути використані для широкого кола задач, що включають як регресію, так і класифікацію, але вони потребують значних обчислювальних ресурсів та великих наборів даних для досягнення високої точності.

Таким чином, вибір між регресією, класифікацією та нейронними мережами залежить від характеру задачі, типу даних, вимог до точності, обчислювальних ресурсів та часу, доступного для навчання моделі. Для задач, де потрібна висока інтерпретованість та швидкість, регресія може бути найбільш підходящим методом. Для складніших задач з багатьма категоріями або нелінійними залежностями класифікація є ефективною. В той час, для найскладніших і найбільш комплексних задач з великими наборами даних, нейронні мережі можуть забезпечити найкращі результати, але їх використання потребує значних ресурсів і часу.

2.7 Висновки до другого розділу

Перш за все, важливо зазначити, що розділ має чітке структурування, яке відображає ключові етапи побудови моделі для аналізу споживання електроенергії. Першим етапом є вибір технік машинного навчання, де розглядаються методи навчання з учителем та без учителя. Це дозволяє зрозуміти, як різні підходи до аналізу даних можуть бути застосовані в енергетичному контексті, в залежності від характеру доступних даних і поставленої задачі. Далі, акцентується увага на використанні нейронних мереж для прогнозування споживання електроенергії. Це є важливим аспектом, оскільки нейронні мережі можуть ефективно працювати з великими обсягами даних і допомагають вирішувати складні задачі, де традиційні методи можуть бути менш ефективними. Таким чином, нейронні мережі забезпечують гнучкість та потужність для моделювання складних

залежностей, що є особливо актуальним в енергетичній сфері. Розділ приділяє увагу методу очищення та підготовки даних, що є необхідним етапом перед застосуванням будь-яких моделей машинного навчання. Відомо, що якість вхідних даних безпосередньо впливає на ефективність побудованої моделі, тому важливість цього етапу важко переоцінити. Підготовка даних включає в себе не тільки очищення від шумів і аномалій, але й правильну організацію даних для подальшого використання в навчанні моделей.

Водночас, у розділі також підкреслюється значущість розуміння проблем і особливостей роботи з великими даними в енергетичній сфері. Великі обсяги даних, що надходять з численних датчиків та систем моніторингу, потребують особливої уваги при виборі методів їх обробки і зберігання. Врахування специфіки таких даних важливе для вибору відповідних інструментів для їх аналізу, що, у свою чергу, впливає на результативність та точність прогностичних моделей.

Ще одним важливим аспектом є аналіз залежності споживання електроенергії від зовнішніх факторів, таких як погодні умови, сезонність та економічні умови. Це дозволяє виявити ключові фактори, що впливають на коливання споживання електроенергії, і забезпечує точність прогнозів у різних умовах. Знання цих залежностей важливе для подальшого планування споживання енергії, оптимізації навантаження на енергетичну систему та зменшення ризиків енергетичних дефіцитів.

Останній підрозділ розділу порівнює різні методи навчання, такі як регресія, класифікація та нейронні мережі. Це порівняння є основою для вибору найефективнішої моделі для вирішення конкретної задачі прогнозування споживання електроенергії. Вибір методу навчання залежить від типу задачі, наявних даних та вимог до точності результатів. У цьому контексті важливою є гнучкість методів машинного навчання, здатність адаптуватися до різних умов і сценаріїв прогнозування.

РОЗДІЛ 3. ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Мета програмної реалізації

Розробити програмний модуль для автоматизованого аналізу споживання енергії, що забезпечить:

- 1) аналіз добових, тижневих і температурних патернів споживання енергії;
- 2) оцінку розподілу споживання між різними секторами (житловий, промисловий, комерційний);
- 3) прогнозування споживання енергії на основі історичних даних із використанням моделей машинного навчання;
- 4) формування звітів із ключовими результатами аналізу та рекомендаціями.

Вхідними даними є CSV-файл з такими полями:

- timestamp – позначка часу;
- hour_of_day – година доби;
- day_of_week – день тижня;
- temperature – температура навколишнього середовища;
- humidity – рівень вологості;
- residential_consumption – споживання в житловому секторі;
- industrial_consumption – споживання в промисловому секторі;
- commercial_consumption – споживання в комерційному секторі;
- total_energy_consumption – загальне споживання енергії.

Перелік завдань для реалізації:

- дослідити середнє споживання енергії залежно від години доби;
- виявити вплив температури на обсяги споживання;
- підрахувати кореляцію між температурою і загальним споживанням;

- обчислити частки енергії, спожитої житловим, промисловим та комерційним секторами;
- візуалізувати розподіл у вигляді діаграми;
- використовувати доступні дані для побудови моделі машинного навчання;
- оцінити точність прогнозу за метриками MSE (mean squared error), R^2 (коефіцієнт детермінації) та MAE (mean absolute error);
- візуалізувати фактичні та прогнозовані значення;
- автоматично створювати текстовий звіт із результатами аналізу;
- зберігати графіки (аналіз патернів, залежність від температури, точність прогнозу тощо) у вигляді зображень.

3.2 Інструменти розробки

Одним із основних інструментів у цьому проекті є бібліотека Pandas, яка використовується для роботи з табличними даними. Вона надає високорівневі структури даних, зокрема DataFrame, що є двовимірною таблицею з мітками рядків і стовпців. Завдяки Pandas можна легко завантажувати дані з CSV-файлів, обробляти їх та проводити агрегацію. В проекті з її допомогою зчитуються вхідні дані, після чого виконується попередня обробка, включаючи розділення за часовими мітками, групування за часовими інтервалами та обчислення середніх значень.

Бібліотека NumPy відіграє ключову роль у чисельних обчисленнях. Її основна структура даних, масиви, оптимізована для швидкої роботи, що особливо важливо при обробці великих обсягів даних. У проекті NumPy застосовується для обчислення статистичних показників, таких як середнє значення, а також для виконання операцій над масивами. Завдяки цьому можна проводити розрахунки з високою продуктивністю.

Для візуалізації даних використовується Matplotlib та її похідна бібліотека Seaborn. Matplotlib є базовим інструментом для створення графіків у Python. Вона надає засоби для побудови лінійних графіків, гістограм, кругових діаграм та багатьох інших типів візуалізації. У проєкті Matplotlib використовується для створення графіків залежності споживання енергії від часу та температури. Seaborn, у свою чергу, надає більш високий рівень абстракції, дозволяючи створювати складні графіки з мінімальною кількістю коду. Наприклад, за допомогою Seaborn було реалізовано побудову графіків важливості ознак, які дозволяють оцінити внесок кожного параметра у прогноз.

Моделювання та прогнозування споживання енергії реалізовано за допомогою бібліотеки scikit-learn, яка є потужним інструментом для машинного навчання. Scikit-learn надає численні алгоритми класифікації, регресії та кластеризації, а також засоби для попередньої обробки даних. У проєкті використовується лінійна регресія, що є простим і водночас ефективним методом для моделювання залежності між набором вхідних змінних та цільовою змінною. Перед застосуванням моделі дані стандартизуються за допомогою StandardScaler, що нормалізує значення до одного масштабу. Це важливо для забезпечення коректної роботи алгоритму, оскільки різниця у масштабах значень може вплинути на точність моделі.

Процес оцінки моделі включає використання метрик, таких як середньоквадратична похибка (MSE), коефіцієнт детермінації (R^2) та середня абсолютна похибка (MAE). Ці метрики дозволяють кількісно оцінити, наскільки добре модель прогнозує значення. MSE визначає середню квадратичну помилку між фактичними та передбаченими значеннями, що дає уявлення про середню помилку в квадраті. R^2 оцінює, яка частина дисперсії у даних пояснюється моделлю, що дозволяє зрозуміти, наскільки добре модель підходить до даних. MAE вимірює середню абсолютну різницю між

фактичними та передбаченими значеннями, що дає змогу оцінити точність моделі з точки зору абсолютних значень.

Крім того, проект включає генерацію звітів, що автоматизує підготовку текстової документації з результатами аналізу. Це досягається завдяки використанню стандартних функцій Python для запису тексту у файли. Генерація звіту дозволяє інтегрувати кількісні результати аналізу, графіки та інтерпретацію отриманих висновків, забезпечуючи повну та зручну для читання документацію.

Особливий акцент у проекті зроблено на модульності та повторюваності. Класова структура програми дозволяє організувати код у логічні блоки, що відповідають за різні аспекти аналізу, включаючи обробку даних, візуалізацію, моделювання та генерацію звітів. Це забезпечує зручність у підтримці та розширенні функціональності.

Реалізація всіх цих компонентів дозволила створити потужний та гнучкий інструмент для аналізу енергоспоживання. Завдяки використанню сучасних бібліотек та методів проект може бути легко адаптований для роботи з іншими наборами даних або для вирішення подібних задач в інших галузях. Поєднання інструментів Pandas, NumPy, Matplotlib, Seaborn та scikit-learn створює комплексну екосистему для аналітики, яка поєднує швидкість обробки даних, точність моделювання та високу якість візуалізації.

3.3 Опис реалізації

Архітектура розробленого програмного коду характеризується чіткою структурою, що забезпечує логічну послідовність виконання, модульність і масштабованість. Основою архітектури є класова структура, яка дозволяє розподілити функціональність між різними методами, кожен з яких відповідає за окремий аспект роботи з даними. Такий підхід сприяє легкому розширенню та підтримці програми. Взаємодія між компонентами коду організована так,

щоб максимально зменшити залежності між модулями, забезпечуючи їх автономність і повторне використання.

Головним елементом архітектури є клас `EnergyConsumptionAnalysis`, що виконує роль центрального об'єкта, навколо якого будується вся логіка програми. Він відповідає за завантаження, зберігання та аналіз даних, а також за формування звітів і візуалізацію результатів. Конструктор класу, реалізований у методі `__init__`, виконує первинне завантаження даних із зовнішнього джерела у форматі CSV, використовуючи бібліотеку `Pandas`. У цьому процесі застосовується механізм обробки винятків, що дозволяє уникнути збоїв у роботі програми в разі проблем із завантаженням файлу. Успішне завантаження підтверджується відповідним повідомленням, тоді як у разі помилки користувач отримує діагностичне повідомлення з деталями.

Після завантаження даних виконується їх первинна обробка, яка включає створення копії оригінальних даних для можливого подальшого аналізу змін. Структура даних, зокрема форматування часових міток, забезпечує можливість виконання аналізу залежно від часових характеристик, таких як година доби або день тижня. Це досягається завдяки інтеграції функціоналу для роботи з датами і часом у бібліотеці `Pandas`, яка забезпечує зручні методи для маніпуляцій із датами.

Логіка виконання комплексного аналізу реалізована в методі `comprehensive_analysis`, який слугує точкою входу до основних аналітичних функцій програми. Цей метод послідовно викликає окремі компоненти аналізу, кожен із яких має чітко визначену задачу. Наприклад, метод `analyze_hourly_consumption` відповідає за побудову графіків середнього споживання енергії залежно від години доби, використовуючи механізми групування даних у `Pandas` і можливості візуалізації `Matplotlib`. Важливим аспектом є збереження результатів аналізу у вигляді графіків, що дозволяє не лише здійснювати аналіз у реальному часі, а й мати доступ до візуалізацій у майбутньому.

Метод `temperature_impact_analysis` аналізує вплив температури на споживання енергії, використовуючи підхід кореляційного аналізу. Він включає створення розсіювального графіка залежності споживання від температури з додаванням лінії тренду, обчисленої за допомогою функцій бібліотеки NumPy. Лінія тренду слугує візуальним показником загальної тенденції, що сприяє більш глибокому розумінню взаємозв'язків у даних. Результати аналізу доповнюються обчисленням коефіцієнта кореляції між змінними, що забезпечує кількісну оцінку сили взаємозв'язку.

Одним із найбільш складних компонентів є метод `advanced_prediction`, який відповідає за прогнозування споживання енергії на основі моделі лінійної регресії. Цей метод включає кілька етапів, таких як вибір вхідних змінних, стандартизація даних, розподіл вибірки на навчальну та тестову, навчання моделі та оцінка її продуктивності. Особливістю реалізації є використання `StandardScaler` для масштабування даних, що дозволяє нормалізувати значення змінних і забезпечує кращу роботу моделі. Після тренування моделі виконуються передбачення для тестової вибірки, а точність оцінюється за допомогою кількох метрик, таких як середньоквадратична похибка (MSE) та коефіцієнт детермінації (R^2). Важливим аспектом є також оцінка важливості вхідних змінних, що візуалізується за допомогою графіка, створеного бібліотекою Seaborn.

Структурний аналіз споживання енергії за секторами реалізований у методі `sector_consumption_analysis`. Він обчислює загальне споживання для кожного сектора, а також їхні частки у загальному обсязі енергоспоживання. Результати цього аналізу представлені у вигляді кругової діаграми, яка забезпечує наочне уявлення про структуру споживання. Такий підхід є корисним для виявлення домінуючих секторів і аналізу їхнього впливу на загальну динаміку енергоспоживання.

Завершальним етапом роботи програми є створення звіту, який автоматично генерується у текстовому файлі. Метод

`generate_comprehensive_report` інтегрує результати всіх етапів аналізу, включаючи статистику по даних, результати прогнозування та візуалізації. Це дозволяє отримати повний огляд виконаної роботи у зручному текстовому форматі, доповненому графічними матеріалами.

Загальна архітектура програми передбачає тісну взаємодію між методами класу, які обмінюються результатами аналізу для досягнення комплексного розуміння даних. Така організація сприяє високій узгодженості компонентів та дозволяє розширювати функціональність без значних змін у коді. Архітектурний дизайн орієнтований на модульність, гнучкість і продуктивність, що робить цей програмний код ефективним рішенням для аналізу великих обсягів енергетичних даних.

3.4 Навчання моделі

Процес роботи розробленого програмного забезпечення охоплює повний цикл аналізу даних, починаючи від їх завантаження та обробки, закінчуючи створенням звітів та візуалізацій, що забезпечує комплексний підхід до вивчення енергетичного споживання. Логіка виконання програми базується на ієрархічній структурі, де кожен компонент відповідає за певну функціональність і взаємодіє з іншими компонентами через стандартизовані інтерфейси. На початковому етапі виконання програми здійснюється завантаження даних, що відбувається за допомогою класу `EnergyConsumptionAnalysis`. Цей клас містить конструктор, який відповідає за ініціалізацію необхідних об'єктів і завантаження вихідного набору даних у вигляді файлу формату CSV. Дані обробляються за допомогою бібліотеки `Pandas`, яка забезпечує ефективне читання файлів і перетворення інформації у вигляді `DataFrame`. Ця структура дозволяє легко маніпулювати даними та виконувати різноманітні операції, такі як групування, фільтрація та агрегація. Після завантаження даних виконується їх первинна перевірка на предмет

відповідності очікуваній структурі, а також обробка можливих винятків, пов'язаних із помилками вмісту файлу або відсутністю необхідних стовпців.

Після успішного завантаження даних виконується підготовка до аналізу, що включає створення копії оригінального набору даних. Це дозволяє зберегти вихідну інформацію незмінною, що є важливим для відтворюваності та можливості повернення до початкових даних у разі потреби. Дані конвертуються у формат, зручний для аналізу, наприклад, часові мітки перетворюються у відповідні типи, що дозволяє виконувати операції, пов'язані з аналізом за годинами, днями чи тижнями.

Ключовою частиною роботи програмного забезпечення є виконання комплексного аналізу даних, що реалізовано в методі `comprehensive_analysis`. Цей метод координує виконання різних етапів аналізу, кожен із яких спрямований на вирішення конкретної задачі. Одним із перших етапів є дослідження патернів споживання енергії залежно від години доби, яке виконується методом `analyze_hourly_consumption`. Тут використовується функціональність `Pandas` для групування даних за годинами, а також бібліотека `Matplotlib` для побудови відповідних графіків. Результати цього етапу дозволяють виявити закономірності у споживанні енергії протягом доби, що може бути корисним для оптимізації графіків роботи енергетичних систем.

Наступним важливим етапом є аналіз впливу температури на споживання енергії, що виконується методом `temperature_impact_analysis`. У цьому випадку дані візуалізуються у вигляді розсіювального графіка, на якому показано залежність споживання енергії від температури. Для більшої наочності до графіка додається лінія тренду, яка обчислюється за допомогою функцій бібліотеки `NumPy`. Лінія тренду дозволяє оцінити загальну тенденцію залежності споживання енергії від температури, а обчислений коефіцієнт кореляції надає кількісну оцінку цього взаємозв'язку.

Особливу увагу приділено прогнозуванню споживання енергії, яке реалізовано в методі `advanced_prediction`. Цей метод включає кілька

послідовних етапів, починаючи з вибору вхідних змінних і стандартизації даних, що виконується за допомогою `StandardScaler`. Це забезпечує нормалізацію значень змінних, що покращує роботу алгоритмів машинного навчання. Після цього дані розбиваються на навчальну і тестову вибірки, що дозволяє оцінити якість моделі на нових даних. Для навчання моделі використовується алгоритм лінійної регресії з бібліотеки `Scikit-learn`, який забезпечує швидку і ефективну побудову передбачувальної моделі. Результати роботи моделі оцінюються за кількома метриками, такими як середньоквадратична похибка, середня абсолютна похибка та коефіцієнт детермінації. Це дозволяє визначити точність і надійність прогнозів.

Одночасно з прогнозуванням виконується оцінка важливості вхідних змінних, що дозволяє виявити ключові фактори, які впливають на споживання енергії. Ця інформація є надзвичайно корисною для прийняття рішень щодо оптимізації роботи енергетичних систем. Результати аналізу важливості змінних візуалізуються у вигляді графіків, які надають чітке уявлення про внесок кожної змінної у модель.

Ще одним значущим компонентом аналізу є дослідження структури споживання енергії за секторами, яке реалізовано в методі `sector_consumption_analysis`. У цьому випадку дані групуються за категоріями, такими як побутове, промислове та комерційне споживання. Обчислюється загальний обсяг споживання для кожного сектора, а також його частка у загальному обсязі енергоспоживання. Результати аналізу подаються у вигляді кругової діаграми, яка дозволяє легко оцінити домінуючі сектори та їхній вплив на загальний рівень енергоспоживання.

Завершальним етапом є створення звіту, який автоматично генерується методом `generate_comprehensive_report`. У звіті узагальнюються результати всіх етапів аналізу, включаючи статистичні показники, графіки та текстові описи. Звіт зберігається у текстовому файлі, що дозволяє користувачеві

ознайомитися з результатами аналізу у зручному форматі. Такий підхід забезпечує прозорість і зрозумілість виконаних розрахунків.

Узагальнюючи, процес роботи програмного забезпечення базується на модульній структурі, що забезпечує чітку послідовність виконання завдань, їхню автономність і взаємодію між компонентами. Завдяки використанню сучасних бібліотек для аналізу даних, машинного навчання та візуалізації забезпечується висока продуктивність і гнучкість програми, що робить її ефективним інструментом для дослідження енергетичного споживання.

3.5 Тестування моделі

На рисунках 3.1-3.2 наведено початкові дані у CSV-файлі та Excel відповідно. Ці дані зчитуються та попередньо оброблюються перед навчанням моделі. Цей процес наведено на рисунку 3.3.

Після навчання моделі було визначено важливість ознак для прогнозування. Це представлено на рисунку 3.4.

На рисунку 3.5 наведено середнє споживання електроенергії за годинами доби.

На рисунку 3.6 представлено передбачені та фактичні значення споживання.

Розподіл споживання електроенергії за секторами наведено на рисунку 3.7. З рисунка 3.7 видно, що промисловий сектор займає найбільше.

```

1 timestamp,hour_of_day,day_of_week,residential_consumption,industrial_consumption,commercial_consumption,temperature,humidity,total_energy_consumption
2 2024-01-01 00:00:00,0,0,11.490142459032699,63.99355436586002,16.62410862512809,10.460962210613046,51.36506395035066,91.26028608597639
3 2024-01-01 01:00:00,1,0,9.585207096486446,59.246336829127685,19.27740664642241,15.69807494610249,59.68796511060916,88.58228754966832
4 2024-01-01 02:00:00,2,0,11.943065614302078,50.59630369920174,16.037900395000193,17.931972332891778,60.18016872044446,77.65811692480233
5 2024-01-01 03:00:00,3,0,14.569089569224076,43.53063222294426,18.4601923518048,29.43843828670095,64.72630345843406,79.69345206877433
6 2024-01-01 04:00:00,4,0,9.297539875829992,56.982233136135896,10.531926665231072,22.78276562266879,46.331416367191736,82.95170097599011
7 2024-01-01 05:00:00,5,0,9.297589129152458,53.934853854217494,21.066468536864694,13.322592154500544,65.92567273847331,87.01029874357171
8 2024-01-01 06:00:00,6,0,14.737638446522174,58.95193220027733,20.006027376811122,22.430181447088184,32.95608368397225,101.30875999794434
9 2024-01-01 07:00:00,7,0,12.302304187458727,56.351718016819696,15.914556844659046,12.263480055308172,53.70115447636189,90.97589059620941
10 2024-01-01 08:00:00,8,0,8.591576842195144,60.49552715319335,23.296228342162024,25.413455269435367,55.117261575484555,95.93301698055383
11 2024-01-01 09:00:00,9,0,11.627680130757895,44.64764788439432,24.68785068813279,17.64437674091061,66.3326826071239,84.98951812687667
12 2024-01-01 10:00:00,10,0,8.609746921562614,63.173940656343255,11.96220067815513,19.531819056304144,60.22630852676869,90.30217985871752
13 2024-01-01 11:00:00,11,0,8.602810739289229,51.975996046923996,16.18637610716893,26.62898323771095,45.944369461437134,80.47174124407847
14 2024-01-01 12:00:00,12,0,10.725886814698102,70.75260872625265,16.15428825106351,13.56418215707875,50.15006874182985,98.6813680762849
15 2024-01-01 13:00:00,13,0,4.260159266026607,43.10812181910432,15.30048404455677,13.014409074435044,61.9868009547407,62.015893791368974
16 2024-01-01 14:00:00,14,0,4.825246502460901,67.3596380316525,24.147374209658743,17.082003368269977,59.20932658135444,95.18755733366899
17 2024-01-01 15:00:00,15,0,8.313137412277083,51.979107834626475,20.039663151331643,26.001308748694147,81.69936588080543,87.19195326159006
18 2024-01-01 16:00:00,16,0,6.961506638996728,43.48581996385552,18.677425793446947,12.403269622615788,55.19166210488021,61.95791408707163
19 2024-01-01 17:00:00,17,0,10.942741997785822,45.16114165945678,9.980688177966233,5.839222009447155,66.96167997159289,66.7972904940317
20 2024-01-01 18:00:00,18,0,7.275927773436367,46.7965269180568,23.17709015104613,17.744205707955008,54.82307427130785,74.103280475809
21 2024-01-01 19:00:00,19,0,5.763088895994126,54.241659464019165,13.803708025732638,22.758703995546,59.119087559709264,71.40319191393083
22 2024-01-01 20:00:00,20,0,14.396946306764661,55.228354880355,20.299663151331643,26.001308748694147,81.69936588080543,87.19195326159006
23 2024-01-01 21:00:00,21,0,9.322671098540393,44.26299996061421,21.386884689790957,17.684193189573005,45.76317363116077,71.76445409881927
24 2024-01-01 22:00:00,22,0,10.202584614063772,49.756454077390174,26.803292979193255,17.94286694419988,41.481572684253415,85.7727498071821
25 2024-01-01 23:00:00,23,0,5.725755441359629,71.42270358611864,13.45589004404439,25.769503671278734,67.83589664798606,88.07820117991724
26 2024-01-02 00:00:00,0,1,8.366851826424451,67.27543170100711,4.902439220895873,10.651291501522897,53.21356938705897,87.14994074312024
27 2024-01-02 01:00:00,1,1,8.332767769129598,54.36323669674032,20.919247723768475,18.057396124084953,56.8176247225738,86.5753257270416
28 2024-01-02 02:00:00,2,1,6.547019267733091,50.380034781682,29.00255907554267,20.95211944962529,52.18843621868886,88.58489251931665
29 2024-01-02 03:00:00,3,1,11.127094055037016,51.200313267182615,26.19473176469719,22.24609360642517,57.40199859232487,93.48021128638287
30 2024-01-02 04:00:00,4,1,6.826867213132299,56.47195939702741,22.527350795833133,15.088529828726603,31.60921340514573,93.24088797835064
31 2024-01-02 05:00:00,5,1,9.12491875062017,39.77207434801595,17.541819071839765,20.13010495374944,20.294921107947985,51.90619178917246,89.75311246195801
32 2024-01-02 06:00:00,6,1,8.19488016331181,47.42623462576655,24.035612983186553,7.558363232935025,68.6223511886977,84.7398913136585
33 2024-01-02 07:00:00,7,1,15.556834553526812,33.314159261224106,15.132269001477656,16.70875793833076,45.837729442073545,73.97822359646867
34 2024-01-02 08:00:00,8,1,9.959508325786198,53.992231226052795,22.381788846781095,22.26754356277234,48.488462634471304,94.8452382464543
35 2024-01-02 09:00:00,9,1,6.826867213132299,56.47195939702741,22.527350795833133,15.088529828726603,31.60921340514573,93.24088797835064
36 2024-01-02 10:00:00,10,1,12.467634736309567,45.16813537700626,25.3010495374944,20.294921107947985,51.90619178917246,89.75311246195801
37 2024-01-02 11:00:00,11,1,6.337469050086933,65.73986763290033,33.79830019455539,22.23436250401349,69.56703186385658,114.9258409695508
38 2024-01-02 12:00:00,12,1,10.626590785014265,37.74234336997659,21.962079835893835,18.287253995711175,65.16946515748371,77.7504160275198
39 2024-01-02 13:00:00,13,1,4.1209896283606735,35.35625119799917,17.455181871462507,20.852318854982855,57.784867370487864,57.39596719556405
40 2024-01-02 14:00:00,14,1,6.015441853047049,52.24451818599972,19.87212843904003,15.186337173576938,59.77699716281947,82.09255508049041

```

Рисунок 3.1 – Початкові дані у CSV-файлі

The screenshot shows the Microsoft Excel application window with the file 'energy_consumption_data.csv' open. The ribbon includes 'Файл', 'Главная', 'Вставка', 'Разметка страницы', 'Формулы', 'Данные', 'Рецензирование', 'Вид', and 'Справка'. The 'Главная' ribbon is active, showing options for font, paragraph, and styles. The spreadsheet grid displays the same data as Figure 3.1, with columns labeled A through Q and rows numbered 1 through 40. The data is organized into columns for timestamp, day of week, residential consumption, industrial consumption, commercial consumption, temperature, humidity, and total energy consumption.

Рисунок 3.2 – Початкові дані в Excel

	timestamp	hour_of_day	day_of_week	residential_consumption	...	commercial_consumption	temperature	humidity	total_energy_consumption
count	1000	1000.000000	1000.000000	1000.000000	...	1000.000000	1000.000000	1000.000000	1000.000000
mean	2024-01-21 19:29:59.999999744	11.436000	2.976000	10.057996	...	20.029171	19.906404	59.507264	82.685474
min	2024-01-01 00:00:00	0.000000	0.000000	0.276198	...	4.902439	5.352757	28.232962	33.716338
25%	2024-01-11 09:45:00	5.000000	1.000000	8.057229	...	16.760002	16.312898	53.173950	73.536467
50%	2024-01-21 19:30:00	11.000000	3.000000	10.075902	...	19.998746	20.000923	59.817580	82.660039
75%	2024-02-01 05:15:00	17.000000	5.000000	11.943832	...	23.304577	23.334727	66.391231	92.378341
max	2024-02-11 15:00:00	23.000000	6.000000	21.558194	...	39.631189	36.215465	91.129102	122.819860
std	Null	6.913009	1.990826	2.937648	...	4.917271	5.135663	9.923802	13.858365

[8 rows x 9 columns]

Рисунок 3.3 – Зчитування та попередня обробка початкових даних

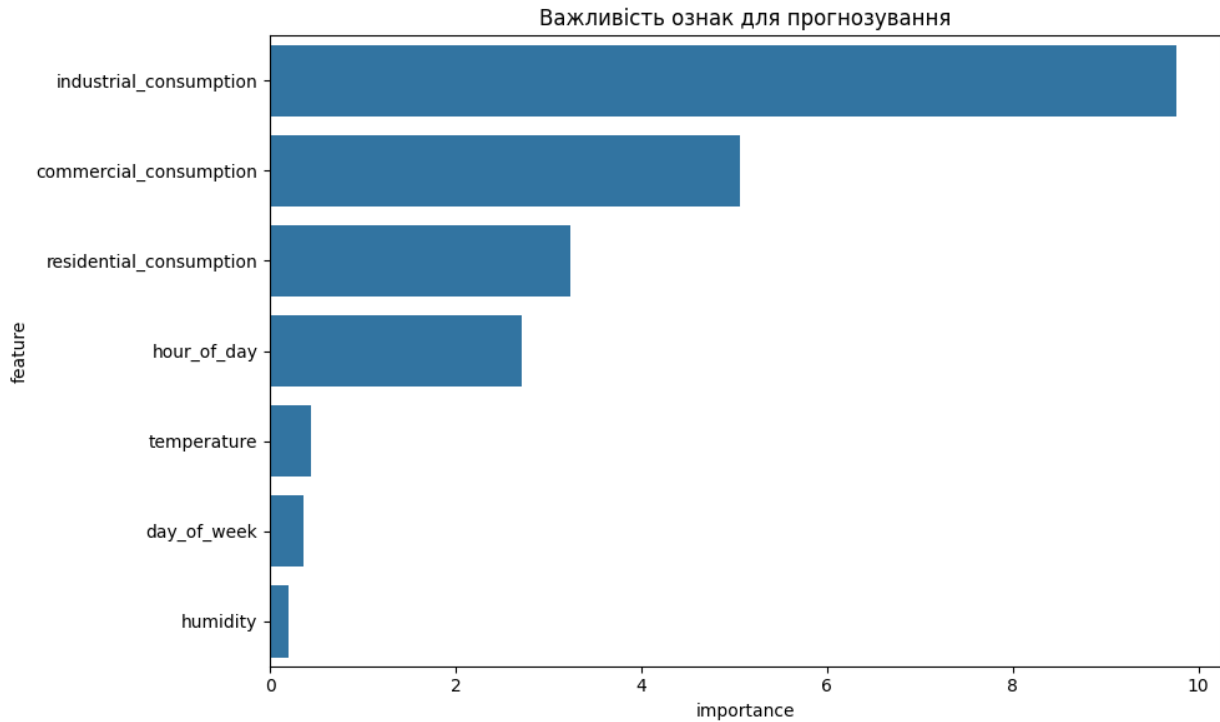


Рисунок 3.4 – Важливість ознак для прогнозування

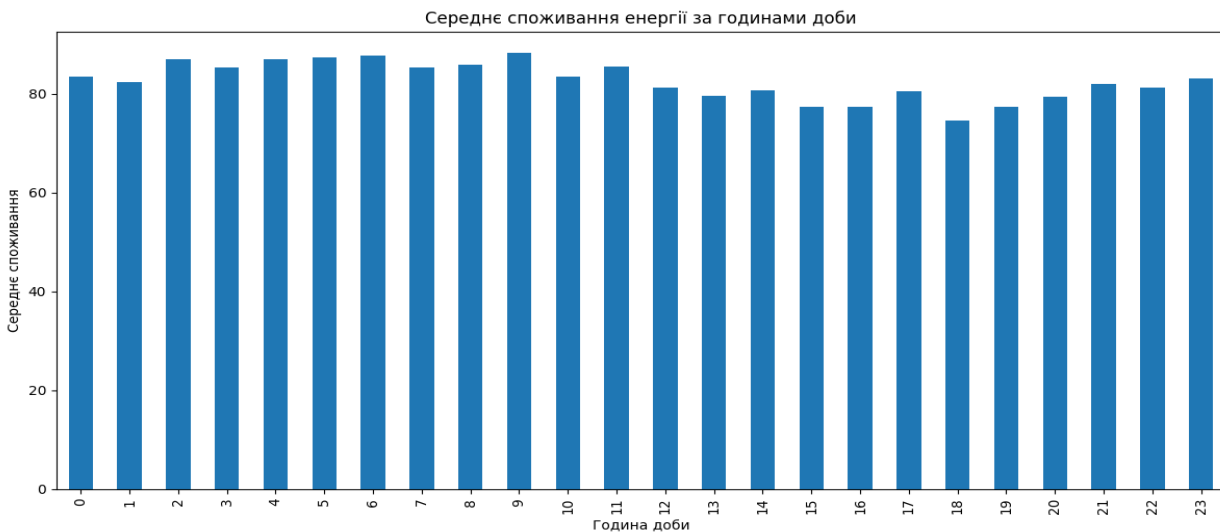


Рисунок 3.5 – Середнє споживання електроенергії за годинами доби

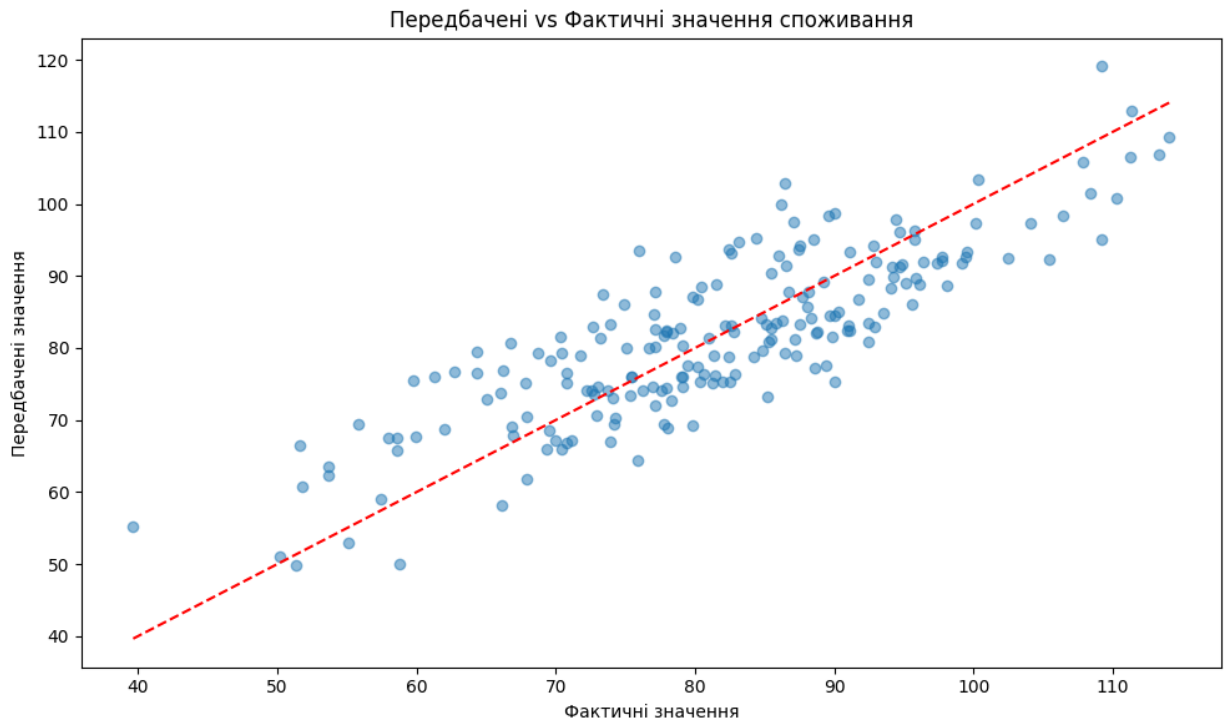


Рисунок 3.6 – Передбачені та фактичні значення споживання

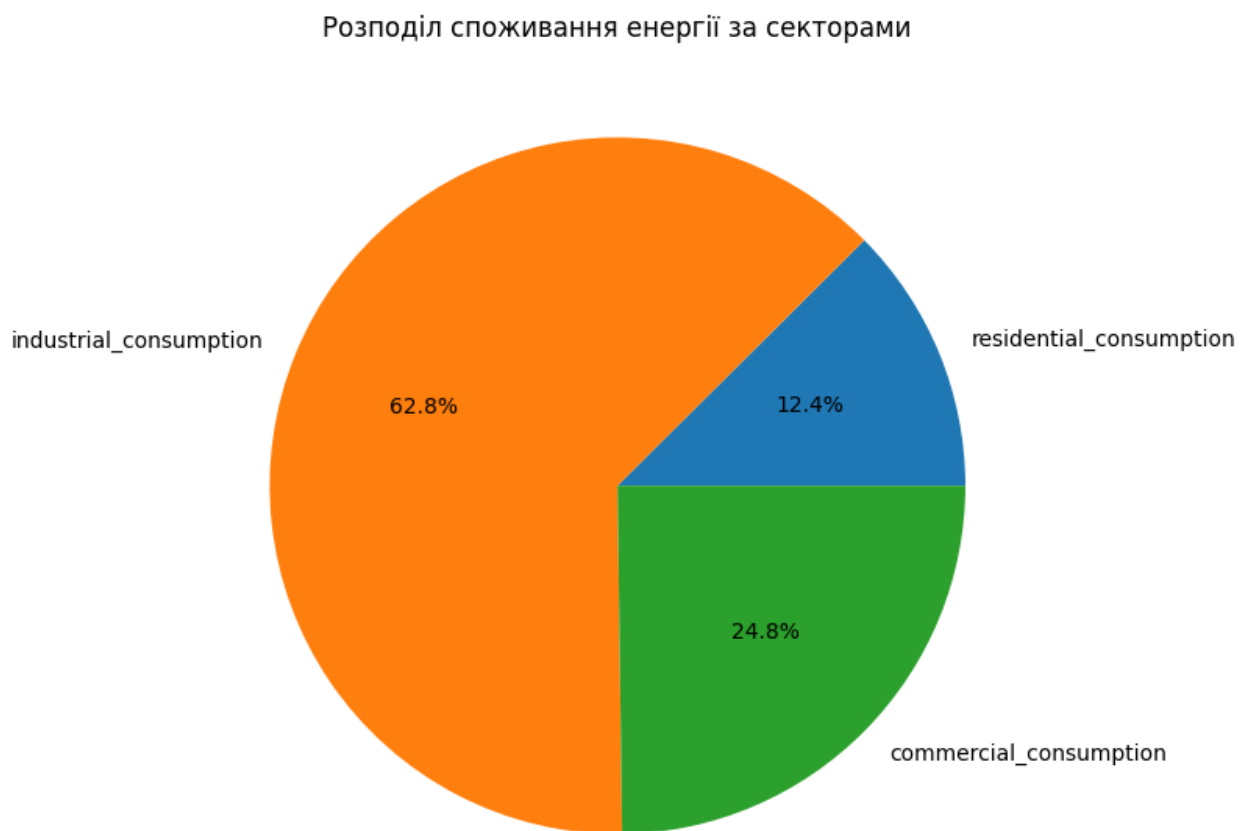


Рисунок 3.7 – Розподіл споживання електроенергії за секторами

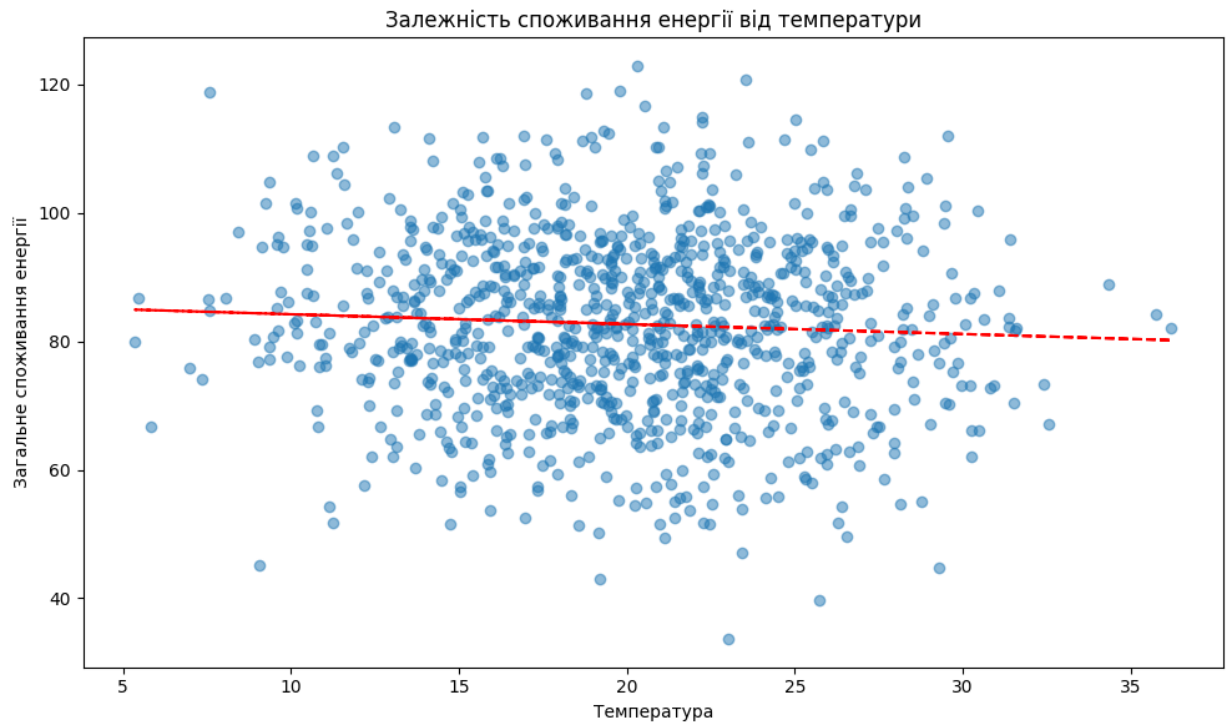


Рисунок 3.8 – Залежність споживання енергії від температури

Загальне споживання енергії:

- середнє значення: 82.69 (умовних одиниць);
- стандартне відхилення: 13.86;
- мінімальне значення: 33.72;
- максимальне значення: 122.82;
- медіана (50%): 82.66.

Інтерпретація:

- значна варіативність споживання ($\text{std} = 13.86$);
- 50% значень знаходяться в діапазоні від 73.54 до 92.38;
- розкид між мінімумом і максимумом становить майже 89 одиниць.

Температура:

- середня температура: 19.91°C;
- діапазон: від 5.35°C до 36.22°C;
- помірні температурні коливання.

Вологість:

- середня вологість: 59.51%;
- діапазон: від 28.23% до 91.13%;
- значні коливання вологості.

Метрики якості моделі:

- MSE (Середньоквадратична похибка), 55.04:
 - показує середню похибку прогнозування;
 - відносно невелике значення;
- R^2 (Коефіцієнт детермінації), 0.70:
 - висока пояснювальна здатність моделі;
 - 70% варіативності даних пояснюється побудованою моделлю;
- MAE (Середня абсолютна похибка), 6.25. У середньому прогноз відхиляється на 6.25 одиниць.

Отже, 3 найважливіших ознаки:

- 1) промислове споживання;
- 2) комерційне споживання;
- 3) житлове споживання.

Менш впливові фактори:

- година доби (2.71);
- температура (0.46);
- вологість (0.21);
- день тижня (0.17).

Отже, після проведення експериментів можна зробити наступні

ВИСНОВКИ:

- 1) модель досить точно прогнозує споживання енергії;
- 2) основний вплив мають обсяги споживання за секторами;
- 3) температура і вологість мають незначний вплив;
- 4) часові характеристики (година, день) також несуттєві.

3.6 Можливість інтеграції з іншими інструментами

Основою інтеграції є використання універсального формату даних у процесах читання та запису інформації. Формат CSV, що застосовується для вхідних і вихідних даних, є загальноприйнятим стандартом у галузі аналітики. Це дозволяє безпосередньо обмінюватися даними між програмою та іншими платформами, такими як SQL-бази даних, інструменти бізнес-аналітики (BI) або навіть прості текстові редактори. Використання бібліотеки Pandas у програмі додає додаткову гнучкість у роботі з іншими форматами, такими як Excel, JSON або навіть спеціалізовані формати, що можуть бути використані в індустрії, наприклад Parquet або HDF5. Це означає, що програмне забезпечення може легко адаптуватися для зчитування даних із будь-яких джерел або збереження результатів у форматі, зручному для подальшого використання іншими системами.

Важливою особливістю є можливість інтеграції програмного забезпечення з реляційними базами даних, такими як PostgreSQL, MySQL або SQLite. Бібліотека SQLAlchemy, яка може бути легко підключена до існуючого коду, дозволяє здійснювати прямий доступ до даних у базах, виконувати складні SQL-запити та зберігати результати аналізу в структурованій формі. Це відкриває можливості для автоматизації процесів у великих корпоративних системах, де дані часто зберігаються у централізованих сховищах. Крім того, така інтеграція може бути корисною для побудови дашбордів або інших інтерфейсів, які вимагають реального часу або періодичного оновлення даних.

У розробці також передбачено можливість інтеграції з хмарними платформами для обчислень і зберігання даних. Завдяки використанню бібліотек, таких як Boto3 для Amazon Web Services (AWS) або Google Cloud Client Libraries, програма може бути розширена для роботи з даними, що зберігаються у хмарних сховищах. Це забезпечує масштабованість, яка є

критичною для роботи з великими обсягами даних або для сценаріїв, що потребують розподілених обчислень. Наприклад, дані можуть зчитуватися безпосередньо з хмарних сховищ, таких як Amazon S3, і використовуватися для аналізу без необхідності попереднього завантаження локально. Результати також можуть бути автоматично завантажені у сховище для подальшого доступу іншими користувачами або програмами.

Ще однією значною можливістю є інтеграція з платформами для побудови машинного навчання та штучного інтелекту. Використання бібліотеки Scikit-learn у поточному коді є першою сходинкою до побудови складніших моделей. Програма може бути легко доповнена функціональністю TensorFlow або PyTorch, що відкриває можливість для інтеграції з сучасними інструментами глибокого навчання. Наприклад, замість використання лінійної регресії, програма може бути адаптована для створення нейронних мереж або складніших моделей прогнозування, які можуть бути розгорнуті як окремі мікросервіси. Ці сервіси можуть бути викликані через RESTful API, що забезпечує доступ до моделей із зовнішніх систем. Важливо зазначити, що програма може бути інтегрована з інструментами візуалізації даних, такими як Tableau, Power BI або Plotly Dash. Генеровані програмою графіки та звіти можуть бути збережені у форматах, сумісних із цими платформами, або передаватися безпосередньо через API. Це дозволяє створювати інтерактивні дашборди, які можуть бути оновлені автоматично на основі нових даних. Для внутрішнього використання програми передбачено функціональність збереження графіків у форматі PNG, але ця функція може бути розширена для підтримки SVG або інтерактивних HTML-візуалізацій. Ще одним напрямом інтеграції є поєднання програми з існуючими системами управління бізнес-процесами або ERP-системами, такими як SAP чи Oracle ERP. Завдяки модульній структурі коду і можливості розширення за допомогою RESTful API або WebSocket, програмне забезпечення може стати частиною більших інформаційних систем, що використовуються у бізнесі для моніторингу і

планування ресурсів. Наприклад, результати аналізу споживання енергії можуть бути автоматично передані в ERP-систему для оптимізації графіків виробництва або прогнозування витрат.

Особливої уваги заслуговує можливість інтеграції з інструментами автоматизації та оркестрації, такими як Apache Airflow або Prefect. Ці платформи можуть бути використані для автоматизації виконання скриптів, наприклад, для запуску аналізу даних за розкладом або на основі певних подій. Завдяки цьому програма може бути легко включена до складних робочих потоків, де вона буде виконувати роль одного з етапів обробки даних. Це може бути корисним для сценаріїв, що вимагають інтеграції різних джерел даних і забезпечення постійного моніторингу стану системи.

У розробці програмного забезпечення також закладено можливість роботи із зовнішніми API для отримання додаткових даних або надсилання результатів. Наприклад, дані про погоду, які можуть впливати на споживання енергії, можуть бути автоматично завантажені через API, такі як OpenWeatherMap або Weatherstack. Це дозволяє автоматизувати процеси і зменшити потребу у ручному введенні інформації.

Програма також може бути інтегрована з інструментами DevOps, такими як Docker або Kubernetes. Використання контейнеризації забезпечує можливість легко розгортати програму у різних середовищах, включаючи локальні сервери, хмарні платформи або гібридні рішення. Контейнеризована версія програми може бути швидко інтегрована у CI/CD-процеси, що спрощує її підтримку і масштабування.

Таким чином, архітектура та функціональність розробленого програмного забезпечення забезпечують широкий спектр можливостей для інтеграції з іншими інструментами і платформами. Це дозволяє використовувати програму не лише як автономний інструмент, але і як частину складніших систем аналізу, прогнозування та управління даними, що

відповідає сучасним вимогам до гнучкості і масштабованості програмного забезпечення.

3.7 Напрямки подальших досліджень

Поточна реалізація базується на класичній лінійній регресії, яка має свої обмеження, особливо коли мова йде про складні та нелінійні залежності між змінними. У подальших дослідженнях доцільно дослідити використання методів машинного навчання, таких як рішення на основі дерев (Decision Trees), бустингові алгоритми (наприклад, Gradient Boosting або XGBoost), а також нейронні мережі, включаючи моделі глибокого навчання. Це дозволить підвищити точність прогнозування, врахувати більшу кількість факторів та адаптувати систему до складніших сценаріїв. Крім того, можливою є інтеграція алгоритмів адаптивного навчання, які можуть оновлювати моделі у реальному часі, використовуючи нові дані.

Ще одним перспективним напрямком є дослідження питань інтерпретованості моделей, що є важливим фактором для їх практичного використання. У сучасному контексті особливої популярності набуває концепція Explainable AI (XAI), яка дозволяє пояснити механізми роботи складних моделей. Інтеграція інструментів для інтерпретації, таких як SHAP (SHapley Additive exPlanations) або LIME (Local Interpretable Model-Agnostic Explanations), може стати важливим кроком у цьому напрямку. Це забезпечить користувачів прозорими поясненнями причин, які лежать в основі прийнятих моделей прогнозування, що особливо важливо у сферах, де аналіз даних має критично важливе значення.

Окремим напрямком досліджень є вдосконалення візуалізації даних. Поточні реалізації графіків та діаграм можуть бути розширені за рахунок інтерактивних візуалізацій, які дозволять користувачам взаємодіяти з даними, змінювати масштаби, обирати різні підмножини та аналізувати дані більш

ефективно. Це може бути досягнуто за допомогою бібліотек Plotly, Dash або D3.js. У перспективі, можлива розробка спеціалізованих дашбордів, які дозволять моніторити ключові показники енергоспоживання у режимі реального часу.

Питання роботи з великими даними також є важливим аспектом для подальших досліджень. Зростаючі обсяги даних вимагають розробки інструментів для їх обробки, зберігання та аналізу у розподілених середовищах. Дослідження у сфері інтеграції з великими платформами, такими як Apache Spark, Hadoop або Dask, дозволять значно підвищити продуктивність програми та забезпечити можливість роботи з мільйонами записів. Також перспективним є використання сучасних баз даних, оптимізованих для роботи з часовими рядами, наприклад, InfluxDB чи TimescaleDB, що дозволить ефективніше зберігати та аналізувати дані про енергоспоживання.

Іншим важливим напрямком є дослідження і розробка функціональності для автоматичного виявлення аномалій. У поточній версії програми немає засобів для автоматичного ідентифікації незвичайних шаблонів у даних, таких як стрибки у споживанні енергії або відхилення, викликані технічними збоями. Використання методів машинного навчання для виявлення аномалій, таких як методи кластеризації, One-Class SVM або Autoencoders, може стати важливим доповненням до функціоналу системи. Така функціональність буде корисною для раннього попередження про можливі проблеми та оптимізації споживання енергії. Крім того, можливим напрямком досліджень є розвиток функціональності для аналізу сценаріїв «що-якщо». Це дозволить користувачам моделювати різні варіанти змін параметрів, таких як температура чи рівень вологості, і аналізувати їх вплив на загальне енергоспоживання. Це може бути реалізовано за допомогою інтерактивних симуляцій або інструментів для моделювання, що дозволять користувачам оцінювати результати потенційних змін у реальному часі.

Не менш важливим напрямком є розвиток системи рекомендацій для оптимізації споживання енергії. На основі даних та прогнозів, програма може пропонувати конкретні дії для зменшення споживання, наприклад, оптимізацію графіків роботи обладнання, зміну налаштувань кліматичних систем або впровадження енергоефективних технологій. Використання методів рекомендаційних систем може значно покращити цінність програми для кінцевих користувачів. Ще одним потенційним напрямком є дослідження питань безпеки та захисту даних. У світі зростаючої кіберзагрози особливу увагу слід приділити безпеці зберігання та обробки даних. Інтеграція засобів шифрування, а також протоколів для забезпечення аутентифікації та авторизації, таких як OAuth або JWT, стане важливим доповненням для захисту конфіденційної інформації.

Дослідження також можуть бути спрямовані на підвищення масштабованості та адаптивності системи. У майбутньому можлива розробка модульної архітектури, яка дозволить підключати нові модулі або функції без необхідності внесення значних змін у кодову базу. Це дозволить програмі адаптуватися до нових викликів і технологій, зберігаючи при цьому високу ефективність.

У світлі розвитку технологій Інтернету речей (IoT), важливим напрямком досліджень є інтеграція з IoT-пристроями. Це дозволить автоматично отримувати дані від розумних лічильників, сенсорів та інших пристроїв, а також забезпечувати управління ними на основі аналізу даних. Це відкриє нові можливості для автоматизації та оптимізації енергоспоживання.

Таким чином, напрямки подальших досліджень у розробленому програмному забезпеченні є надзвичайно широкими та охоплюють як технічні аспекти, так і прикладні задачі. Їх реалізація дозволить значно підвищити ефективність, точність і функціональність програми, забезпечуючи її відповідність сучасним потребам і викликам у сфері аналізу даних та енергоспоживання.

3.8 Висновки до третього розділу

У розділі представлено мету програмної реалізації, яка задає основний напрям роботи. Це дозволяє визначити, які саме завдання вирішує програмне забезпечення та яких цілей воно повинно досягти. Детально описані інструменти розробки, що вказує на використання сучасних технологій, які забезпечують ефективність розробки та можливість інтеграції з іншими системами. Надається опис реалізації, що дозволяє зрозуміти основну структуру програмного забезпечення, включаючи його ключові компоненти та їх взаємодію. Це дає змогу оцінити логіку розробки та її відповідність поставленій меті. Підрозділ 3.4 присвячено процесу навчання моделі, що підкреслює використання алгоритмів машинного навчання для виконання поставлених завдань. Увага до цього етапу свідчить про важливість налаштування моделі для досягнення високої точності прогнозування.

Розглядається процес тестування моделі, що вказує на підхід до оцінювання її ефективності, виявлення недоліків і забезпечення її відповідності вимогам до якості. Розглядається можливість інтеграції з іншими інструментами, що свідчить про гнучкість програмного забезпечення і його придатність для використання в ширшому контексті. Це забезпечує його адаптивність до потреб користувачів та інших систем.

Аналізуються можливі напрямки подальших досліджень, акцентуючи увагу на перспективах удосконалення та розширення функціональності програми. Це свідчить про стратегічний підхід до розвитку програмного продукту.

Загалом, розділ демонструє систематичний підхід до розробки програмного забезпечення, акцентуючи увагу на його структурі, функціональності та потенційних можливостях. Він підтверджує, що програмна реалізація була виконана із застосуванням сучасних технологій та відповідає науковим і практичним вимогам.

ВИСНОВКИ

Робота розглядає широкий спектр задач і методів, пов'язаних із застосуванням машинного навчання для аналізу і прогнозування даних споживання електроенергії. У дослідженні висвітлено сучасні підходи до обробки великих обсягів енергетичних даних, зокрема статистичні методи, класичні підходи до прогнозування та інноваційні алгоритми машинного навчання. Увага приділена питанням оптимізації прогнозування споживання електроенергії, виявлення аномалій, обробки даних, а також врахування зовнішніх факторів, таких як погода, сезонність і соціально-економічні зміни.

Аналіз предметної області засвідчив, що застосування машинного навчання дозволяє значно підвищити точність прогнозування, особливо в умовах високої варіативності даних і впливу численних факторів. Алгоритми нейронних мереж, глибокого навчання, регресії, класифікації та кластеризації продемонстрували здатність враховувати складні нелінійні залежності і створювати адаптивні моделі. Зокрема, впровадження нейронних мереж, таких як рекурентні моделі та LSTM, дозволяє ефективно працювати з часовими рядами і довгостроковими залежностями, що є критично важливим для прогнозування попиту на електроенергію.

Важливим аспектом стало дослідження методів очищення та підготовки даних, що є основою для коректної роботи моделей. Виявлення і корекція пропущених, аномальних чи некоректних значень, а також нормалізація і трансформація даних дозволяють забезпечити необхідну якість вхідної інформації. Це, у свою чергу, впливає на точність прогнозів і загальну ефективність систем.

Практичне застосування результатів дослідження включає можливість розробки розумних енергетичних мереж, оптимізації енергетичних витрат і підвищення стабільності систем. Використання алгоритмів для виявлення

аномалій дозволяє оперативно реагувати на відхилення у споживанні, попереджати несправності і знижувати ризики аварійних ситуацій.

Наукова новизна роботи полягає в розробці комбінованих методів, які враховують широкий спектр впливових факторів і адаптуються до змінних умов. Запропоновані підходи дозволяють інтегрувати машинне навчання в сучасні енергетичні системи, сприяючи їхній ефективності, надійності та екологічності. Окрім цього, дослідження створює основу для подальшого розвитку теоретичних і практичних аспектів використання штучного інтелекту в енергетиці.

Підсумовуючи, кваліфікаційна робота магістра вносить значний вклад у розвиток сучасних підходів до аналізу енергетичних даних, забезпечуючи інтеграцію машинного навчання в практику управління енергетичними ресурсами. Отримані результати мають важливе значення для енергетичних компаній, дослідників і розробників, які працюють над створенням стійких і ефективних енергетичних систем.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. P. Wang, W. Zhang, M. Ma, W. Jin, and X. Wang, “Research on load data classification method based on the characteristics of electricity consumption behavior by extreme learning machine”, in *2023 IEEE Int. Conf. Sensors, Electron. Comput. Eng. (ICSECE)*, Jinzhou, China, Aug. 18–20, 2023. IEEE, 2023. DOI: <https://doi.org/10.1109/icsece58870.2023.10263477>
2. V. P. P. B. Badiger, P. N. Kulkarni, P. R. Prabhu, P. V. Shah, and S. V. B., “A Novel Approach to Design and Develop A Learning Based Framework for Electricity Consumption”, in *2021 IEEE Int. Conf. Mobile Netw. Wireless Commun. (ICMNWC)*, Tumkur, Karnataka, India, Dec. 3–4, 2021. IEEE, 2021. DOI: <https://doi.org/10.1109/icmnwc52512.2021.9688343>
3. J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, and C. K. Lebekwe, “Electricity Load Prediction Using Machine Learning”, in *2022 Int. Conf. Smart Appl., Commun. Netw. (SmartNets)*, Palapye, Botswana, Nov. 29–Dec. 1, 2022. IEEE, 2022. DOI: <https://doi.org/10.1109/smartnets55823.2022.9993990>
4. V. Raviprabhakaran, P. Pranay, B. Nendralla, and L. S. Pranay, “Household Power Consumption Analysis using Machine Learning”, in *2024 IEEE 4th Int. Conf. Sustain. Energy Future Electric Transp. (SEFET)*, Hyderabad, India, Jul. 31–Aug. 3, 2024. IEEE, 2024, pp. 1–6. DOI: <https://doi.org/10.1109/sefet61574.2024.10718254>
5. I. U. Khan, N. Javeid, C. J. Taylor, K. A. A. Gamage, and X. Ma, “A Stacked Machine and Deep Learning-Based Approach for Analysing Electricity Theft in Smart Grids”, *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 1633–1644, Mar. 2022. DOI: <https://doi.org/10.1109/tsg.2021.3134018>
6. P. Manandhar, M. Lin, J. D. B. Ramirez, and E. Rodriguez-Ubinas, “Using Occupant's Data for Electricity Load Prediction based on Machine Learning”, in *2021 2nd Int. Conf. Electron., Commun. Inf. Technol. (CECIT)*, Sanya,

China, Dec. 27–29, 2021. IEEE, 2021. DOI: <https://doi.org/10.1109/cecit53797.2021.00225>

7. M. Ahammad and D. M. Farid, “Electricity Theft Detection using Binary-Class Imbalanced Classification”, in *2023 26th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Cox's Bazar, Bangladesh, Dec. 13–15, 2023. IEEE, 2023. DOI: <https://doi.org/10.1109/iccit60459.2023.10441636>

8. K. Haridass, L. C, M. G, P. Madhavan, and R. Durai, “Electricity Consumption Data Analysis Using Spark Framework”, in *2022 1st Int. Conf. Comput. Sci. Technol. (ICCST)*, CHENNAI, India, Nov. 9–10, 2022. IEEE, 2022. DOI: <https://doi.org/10.1109/iccst55948.2022.10040282>

9. D. I. Jurj, D. D. Micu, L. Czumbil, A. G. Berciu, M. Lancrajan, and D. M. Barar, “Analysis of Data Cleaning Techniques for Electrical Energy Consumption of a Public Building”, in *2020 55th Int. Univ. Power Eng. Conf. (UPEC)*, Torino, Italy, Sep. 1–4, 2020. IEEE, 2020. DOI: <https://doi.org/10.1109/upec49904.2020.9209781>

10. H. Qin, Z. Zhao, X. Cao, S. Li, J. Zhang, and Z. Yang, “Building Electricity Consumption Analysis: Data-Driven Approach with Preprocessing, Visualization, and Cluster Analysis”, in *2023 Int. Conf. Electron. Devices, Comput. Sci. (ICEDCS)*, Marseille, France, Sep. 22–24, 2023. IEEE, 2023. DOI: <https://doi.org/10.1109/icedcs60513.2023.00015>

11. X. Xiong and Y. Wei, “The Analysis and Predication of Energy Use in Smart Homes Based on Machine Learning”, in *2020 Int. Conf. Comput. Data Sci. (CDS)*, Stanford, CA, Aug. 1–2, 2020. IEEE, 2020. DOI: <https://doi.org/10.1109/cds49703.2020.00080>

12. X. Zhou, Z. Xu, X. Yu, and Y. Xia, “The Analysis and Prediction of Power Consumption in Industry Based on Machine Learning”, in *2019 Chin. Automat. Congr. (CAC)*, Hangzhou, China, Nov. 22–24, 2019. IEEE, 2019. DOI: <https://doi.org/10.1109/cac48633.2019.8996219>

13. N. Chauhan, S. Bansal, and M. Malik, “A Comparative Analysis for Short and Long-Term Time Series for Energy Consumption Predictions”, in *2024 Int. Conf. Comput. Intell. Comput. Appl. (ICCICA)*, Samalkha, India, May 23–24, 2024. IEEE, 2024. DOI: <https://doi.org/10.1109/iccica60014.2024.10584996>

14. M. Wasesa *et al.*, “Predicting Electricity Consumption in Microgrid-Based Educational Building Using Google Trends, Google Mobility, and COVID-19 Data in the Context of COVID-19 Pandemic”, *IEEE Access*, vol. 10, pp. 32255–32270, 2022. DOI: <https://doi.org/10.1109/access.2022.3161654>

ДОДАТКИ

Додаток А

Можливість генерації даних за їх відсутності

```
import pandas as pd
import numpy as np
from datetime import datetime, timedelta

def generate_energy_consumption_data(num_records=1000):
    """
    Генерація синтетичних даних про споживання енергії

    Args:
        num_records (int): Кількість записів для генерації

    Returns:
        pandas.DataFrame: DataFrame з даними про споживання енергії
    """

    np.random.seed(42)

    start_date = datetime(2024, 1, 1)
    dates = [start_date + timedelta(hours=i) for i in range(num_records)]

    data = {
        'timestamp': dates,
```

```
'hour_of_day': [date.hour for date in dates],
'day_of_week': [date.weekday() for date in dates],

'residential_consumption': np.random.normal(10, 3, num_records),
'industrial_consumption': np.random.normal(50, 10, num_records),
'commercial_consumption': np.random.normal(20, 5, num_records),

'temperature': np.random.normal(20, 5, num_records),
'humidity': np.random.normal(60, 10, num_records),

'total_energy_consumption': 0
}

df = pd.DataFrame(data)

df['total_energy_consumption'] = (
    df['residential_consumption'] +
    df['industrial_consumption'] +
    df['commercial_consumption'] +
    np.random.normal(0, 2, num_records)
)

df['total_energy_consumption'] += np.sin(df['hour_of_day'] * np.pi / 12) * 5
```

```

df['total_energy_consumption'] += np.sin(df.index * np.pi / 365) * 10

return df

def save_energy_data(df, filename='energy_consumption_data.csv'):
    """
    Збереження згенерованих даних у CSV файл

    Args:
        df (pandas.DataFrame): DataFrame з даними
        filename (str): Назва файлу для збереження
    """
    df.to_csv(filename, index=False)
    print(f"Дані збережено у файл: {filename}")

    print("\nОгляд згенерованих даних:")
    print(df.describe())

def main():

    energy_data = generate_energy_consumption_data(num_records=1000)

    save_energy_data(energy_data)

if __name__ == "__main__":
    main()

```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

class EnergyConsumptionAnalysis:
    def __init__(self, file_path):
        try:
            self.data = pd.read_csv(file_path, parse_dates=['timestamp'])
            self.original_data = self.data.copy()
            print("Дані успішно завантажені")
        except Exception as e:
            print(f"Помилка завантаження даних: {e}")
            self.data = None

    def comprehensive_analysis(self):
        """
        Комплексний аналіз споживання енергії з детальною звітністю
        """
        if self.data is not None:
            self.analyze_hourly_consumption()
```



```
self.temperature_impact_analysis()

prediction_results = self.advanced_prediction()

self.sector_consumption_analysis()

self.generate_comprehensive_report(prediction_results)

def analyze_hourly_consumption(self):
    """
    Аналіз споживання енергії за годинами доби
    """
    plt.figure(figsize=(12, 6))
    hourly_consumption =
self.data.groupby('hour_of_day')['total_energy_consumption'].mean()
    hourly_consumption.plot(kind='bar')
    plt.title('Середнє споживання енергії за годинами доби')
    plt.xlabel('Година доби')
    plt.ylabel('Середнє споживання')
    plt.tight_layout()
    plt.savefig('hourly_consumption.png')
    plt.close()

def temperature_impact_analysis(self):
```

```

"""
Аналіз впливу температури на споживання енергії
"""

plt.figure(figsize=(10, 6))
plt.scatter(self.data['temperature'], self.data['total_energy_consumption'],
alpha=0.5)
plt.title('Залежність споживання енергії від температури')
plt.xlabel('Температура')
plt.ylabel('Загальне споживання енергії')

z = np.polyfit(self.data['temperature'], self.data['total_energy_consumption'],
1)
p = np.poly1d(z)
plt.plot(self.data['temperature'], p(self.data['temperature']), "r--")

plt.tight_layout()
plt.savefig('temperature_impact.png')
plt.close()

correlation =
self.data['temperature'].corr(self.data['total_energy_consumption'])
return correlation

def advanced_prediction(self):
"""
Розширене прогнозування споживання з детальним аналізом
"""

```

```
input_columns = [  
    'hour_of_day',  
    'day_of_week',  
    'temperature',  
    'humidity',  
    'residential_consumption',  
    'industrial_consumption',  
    'commercial_consumption'  
]  
  
X = self.data[input_columns]  
y = self.data['total_energy_consumption']  
  
scaler = StandardScaler()  
X_scaled = scaler.fit_transform(X)  
  
X_train, X_test, y_train, y_test = train_test_split(  
    X_scaled, y, test_size=0.2, random_state=42  
)  
  
model = LinearRegression()  
model.fit(X_train, y_train)  
  
y_pred = model.predict(X_test)  
  
mse = mean_squared_error(y_test, y_pred)  
r2 = r2_score(y_test, y_pred)  
mae = np.mean(np.abs(y_test - y_pred))
```

```
plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred, alpha=0.5)
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'r--')
plt.title('Передбачені vs Фактичні значення споживання')
plt.xlabel('Фактичні значення')
plt.ylabel('Передбачені значення')
plt.tight_layout()
plt.savefig('prediction_scatter.png')
plt.close()
```

```
feature_importance = pd.DataFrame({
    'feature': input_columns,
    'importance': np.abs(model.coef_)
}).sort_values('importance', ascending=False)
```

```
plt.figure(figsize=(10, 6))
sns.barplot(x='importance', y='feature', data=feature_importance)
plt.title('Важливість ознак для прогнозування')
plt.tight_layout()
plt.savefig('feature_importance.png')
plt.close()
```

```
return {
    'mse': mse,
    'r2': r2,
    'mae': mae,
    'feature_importance': feature_importance
```

```

    }

def sector_consumption_analysis(self):
    """
    Аналіз споживання за секторами
    """

    sectors = ['residential_consumption', 'industrial_consumption',
'commercial_consumption']
    sector_totals = self.data[sectors].sum()
    sector_percentages = sector_totals / sector_totals.sum() * 100

    plt.figure(figsize=(8, 8))
    plt.pie(sector_percentages, labels=sectors, autopct='% 1.1f%%')
    plt.title('Розподіл споживання енергії за секторами')
    plt.tight_layout()
    plt.savefig('sector_consumption_pie.png')
    plt.close()

    return sector_percentages

def generate_comprehensive_report(self, prediction_results):
    """
    Створення детального звіту з результатами аналізу
    """

    with open('energy_consumption_detailed_report.txt', 'w', encoding='utf-8') as
f:
        f.write("КОМПЛЕКСНИЙ ЗВІТ АНАЛІЗУ СПОЖИВАННЯ
ЕНЕРГІЇ\n\n")

```

```
f.write("1. ЗАГАЛЬНА СТАТИСТИКА:\n")
f.write(str(self.data[['total_energy_consumption', 'temperature',
'humidity']].describe()) + "\n\n")

f.write("2. РЕЗУЛЬТАТИ ПРОГНОЗУВАННЯ:\n")
f.write(f" - Середньоквадратична похибка (MSE):
{prediction_results['mse']:.2f}\n")
f.write(f" - Коефіцієнт детермінації (R2):
{prediction_results['r2']:.2f}\n")
f.write(f" - Середня абсолютна похибка (MAE):
{prediction_results['mae']:.2f}\n\n")

f.write("3. ВАЖЛИВІСТЬ ОЗНАК ДЛЯ ПРОГНОЗУВАННЯ:\n")
f.write(str(prediction_results['feature_importance']) + "\n")

def main():
    FILE_PATH = 'energy_consumption_data.csv'

    analyzer = EnergyConsumptionAnalysis(FILE_PATH)
    analyzer.comprehensive_analysis()

if __name__ == "__main__":
    main()
```