

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

Факультет комп'ютерно-інформаційних систем і програмної інженерії

(повна назва факультету)

Кафедра комп'ютерних наук

(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

бакалавр

(назва освітнього ступеня)

на тему: Розробка програмного забезпечення для
прогнозу енергоспоживання засобами машинного навчання

Виконав(ла): студент(ка) 4 курсу, групи СН-42

спеціальності 122 Комп'ютерні науки

(шифр і назва спеціальності)

Шеремета В.В.
(підпис) (прізвище та ініціали)

Керівник Гащин Н.Б.
(підпис) (прізвище та ініціали)

Нормоконтроль Боднарчук І.О.
(підпис) (прізвище та ініціали)

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

Рецензент Боднарчук І.О.
(підпис) (прізвище та ініціали)

Тернопіль
2024

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

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

Кафедра комп'ютерних наук
(повна назва кафедри)

ЗАТВЕРДЖУЮ

Завідувач кафедри

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

«27» червня 2024 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня бакалавр
(назва освітнього ступеня)

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

студенту Шеремета Вадим Васильович
(прізвище, ім'я, по батькові)

1. Тема роботи Розробка програмного забезпечення для прогнозу енергоспоживання засобами машинного навчання

Керівник роботи к.т.н., доц. Гащин Н.Б.
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «29» квітня 2024 року № 4/7-470

2. Термін подання студентом завершеної роботи 27 червня 2024 р.

3. Вихідні дані до роботи Літературні джерела з тематики роботи

4. Зміст роботи (перелік питань, які потрібно розробити)

ВСТУП; 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ; 1.1 Аналіз глобального попиту на електроенергію 1.2 Країни, що розвиваються, є двигунами зростання глобального попиту на електроенергію 1.3 Перспективи на швидке відновлення енергоємних галузей у Європейському Союзі 1.4 Прогноз постачання чистої електроенергії 1.5 Перспективи виробництва ядерної енергії 2 ОГЛЯД АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ЕНЕРГОСПОЖИВАННЯ 2.1 Принципи Використання Automl 3 ПРАКТИЧНА ЧАСТИНА РОБОТИ 3.1 Основні технології front-end розробки 3.2 Аналіз вхідного набору даних 3.3 Тренування та оцінка моделі прогнозування Prophet 3.4 Перетворюйте дані часових рядів у табличні дані за допомогою функції 3.5 Навчання та оцінка моделі GradientBoostingClassifier на основі представлених табличних даних 3.6 Використання AutoML для оптимізації 4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ; ВИСНОВКИ; СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ; ДОДАТКИ

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Безпека життєдіяльності, основи охорони праці	Сенчишин В. С., доцент кафедри МТ		

7. Дата видачі завдання 29 січня 2024 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	24.01.24-27.01.24	Виконано
2.	Підбір джерел по темі роботи	28.01.24 – 01.04.24	Виконано
3.	Оформлення першого розділу	15.04.2024	Виконано
4.	Оформлення другого розділу	20.04.2024	
5.	Оформлення третього розділу	30.04.2024	Виконано
6.	Виконання завдання до підрозділу «Безпека		
7.	життєдіяльності, основи охорони праці»	15.05.2024	Виконано
8.	Оформлення кваліфікаційної роботи	07.06.2024	Виконано
9.	Перевірка на плагіат	07.06.2024	Виконано
10.	Нормоконтроль	09.06.2024	Виконано
11.	Попередній захист кваліфікаційної роботи	11.06.2024	Виконано
12.	Захист кваліфікаційної роботи	28.06.2024	
13.			
14.			
15.			

Студент

_____ (підпис)

Шеремета В.В.

_____ (прізвище та ініціали)

Керівник роботи

_____ (підпис)

Гащин Н.Б.

_____ (прізвище та ініціали)

АНОТАЦІЯ

"Розробка програмного забезпечення для прогнозу енергоспоживання засобами машинного навчання" // Кваліфікаційна робота освітнього рівня «Бакалавр» // Шеремета Вадим Васильович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних наук, група СН-42 // Тернопіль, 2024 // с. – 65 , рис. – 20, джерел – 18, додатків – 1, сторінок додатків – 8.

Ключові слова: прогнозування, енергоспоживання, машинне навчання, алгоритми навчання

Прогнозування обсягів енергоспоживання є критично важливим для України в сучасних умовах, зокрема через військову агресію росії. Цей фактор додає до існуючих викликів, таких як економічна нестабільність, енергетична залежність, кліматичні зміни та необхідність впровадження ефективних енергетичних політик. У сучасному глобалізованому світі, де енергетичні ресурси відіграють ключову роль у розвитку економіки, забезпечення стабільного та прогнозованого енергоспоживання стає основою енергетичної безпеки країни.

Агресія Росії проти України створила нові загрози для енергетичної інфраструктури та підвищила важливість точного прогнозування обсягів енергоспоживання. В умовах військових дій та можливих диверсій на об'єктах критичної інфраструктури, здатність передбачати та швидко реагувати на зміни в енергоспоживанні є життєво важливою. Прогнозування дозволяє уряду та підприємствам планувати свої дії, що знижує ризики неочікуваних збоїв у постачанні енергії.

Енергетична залежність України від імпорту газу та нафти, значна частина яких надходила з росії, підкреслює необхідність розвитку власних

енергоресурсів та зменшення залежності від зовнішніх постачальників. Прогнозування допомагає визначити оптимальні шляхи забезпечення енергетичної незалежності та ефективного використання наявних ресурсів.

Кліматичні зміни та зобов'язання України щодо зниження викидів парникових газів відповідно до міжнародних угод також вимагають ретельного прогнозування обсягів енергоспоживання. Зменшення викидів можливо лише за умови оптимізації енергоспоживання та переходу на відновлювані джерела енергії. Прогнозування допомагає визначити найбільш ефективні шляхи інтеграції зелених технологій у енергетичний сектор.

Важливою складовою прогнозування є врахування соціально-економічних тенденцій та демографічних змін, що впливають на попит на енергію. Це дозволяє розробляти довгострокові стратегії розвитку енергетичного сектору, спрямовані на підвищення енергоефективності та стабільності енергозабезпечення.

Таким чином, прогнозування обсягів енергоспоживання є ключовим елементом забезпечення енергетичної безпеки України. Воно сприяє економічній стабільності, зменшенню залежності від імпорту енергоносіїв, виконанню міжнародних екологічних зобов'язань та оптимізації внутрішнього енергоспоживання. Ефективне прогнозування дозволяє створити надійну основу для сталого розвитку енергетичного сектору країни, особливо в умовах агресії росії.

ANNOTATION

"Development of Software for Forecasting Energy Consumption by Means of Machine Learning" // Qualification work of the educational level "Bachelor" // Sheremeta Vadym // Ternopil Ivan Puluj National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science, Group CH-42 // Ternopil, 2024 // p. – 65, fig. – 20, references – 18, annexes – 1, pages for annexes – 8.

Keywords: forecasting, energy consumption, machine learning, learning algorithms

Forecasting energy consumption is critically important for Ukraine in modern conditions, especially considering Russia's military aggression. This factor adds to existing challenges such as economic instability, energy dependence, climate change, and the necessity to implement effective energy policies. In the modern globalized world, where energy resources play a key role in economic development, ensuring stable and predictable energy consumption becomes the foundation of the country's energy security.

Russia's aggression against Ukraine has created new threats to the energy infrastructure and increased the importance of accurate energy consumption forecasting. In conditions of military actions and possible sabotage of critical infrastructure facilities, the ability to predict and quickly respond to changes in energy consumption is vital. Forecasting allows the government and enterprises to plan their actions, reducing the risks of unexpected energy supply disruptions.

Ukraine's energy dependence on imported gas and oil, a significant portion of which previously came from Russia, underscores the need to develop domestic energy resources and reduce reliance on external suppliers. Forecasting helps identify optimal ways to ensure energy independence and the efficient use of available resources.

Climate change and Ukraine's commitments to reduce greenhouse gas emissions according to international agreements also require careful forecasting of energy consumption. Emission reduction is possible only with optimized energy consumption and a transition to renewable energy sources. Forecasting helps determine the most effective ways to integrate green technologies into the energy sector.

An important component of forecasting is considering socio-economic trends and demographic changes that affect energy demand. This enables the development of long-term strategies for the energy sector aimed at increasing energy efficiency and the stability of energy supply.

Thus, forecasting energy consumption is a key element of ensuring Ukraine's energy security. It promotes economic stability, reduces dependence on imported energy resources, fulfills international environmental obligations, and optimizes domestic energy consumption. Effective forecasting creates a reliable foundation for the sustainable development of the country's energy sector, especially under conditions of Russian aggression.

ЗМІСТ

ВСТУП.....	9
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	11
1.1 Аналіз глобального попиту на електроенергію.....	11
1.2 Країни, що розвиваються, є двигунами зростання глобального попиту на електроенергію.....	12
1.3 Перспективи на швидке відновлення енергоємних галузей у Європейському Союзі.....	14
1.4 Прогноз постачання чистої електроенергії.....	16
1.5 Перспективи виробництва ядерної енергії.....	19
2 ОГЛЯД АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ЕНЕРГОСПОЖИВАННЯ.....	27
2.1 Принципи Використання AutoML.....	27
2.1.1 Огляд AutoML.....	28
2.1.2 Переваги AutoML у машинному навчанні.....	29
2.1.3 Недоліки AutoML.....	30
2.1.4 Практичне застосування автоматизації машинного навчання.....	32
2.2 Алгоритм Gradient Boosting для задач прогнозу енергоспоживання.....	33
2.2.1 Приклад алгоритму gradient boosting (посилення градієнта).....	35
2.2.2 Впровадження gbm за допомогою scikit-learn.....	39
3 ПРАКТИЧНА ЧАСТИНА РОБОТИ.....	42
3.1 Загальний опис роботи програми.....	42
3.2 Аналіз вхідного набору даних.....	43
3.3 Тренування та оцінка моделі прогнозування Prophet.....	44
3.4 Перетворюйте дані часових рядів у табличні дані за допомогою функції.....	46
3.5 Навчання та оцінка моделі GradientBoostingClassifier на основі представлених табличних даних.....	49
3.6 Використання AutoML для оптимізації.....	50
4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ.....	53

4.1 Охорона праці та її актуальність в іт-сфері	53
4.2 Шкідлива дія шуму та вібрації і захист від неї.....	57
ВИСНОВКИ.....	62
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	63
ДОДАТКИ	

ВСТУП

Наукові дослідження в галузі прогнозування енергоспоживання є актуальними та багатограними, оскільки ця проблема має ключове значення для забезпечення сталого розвитку та енергетичної безпеки. Сучасні дослідження зосереджені на розробці точних моделей прогнозування, що використовують новітні досягнення в галузі машинного навчання та великих даних.

Методи машинного навчання, такі як нейронні мережі, дерева рішень, та ансамблеві методи, активно застосовуються для аналізу великих обсягів історичних даних про енергоспоживання. Ці алгоритми дозволяють виявити складні закономірності та тренди, які можуть бути використані для більш точного прогнозування.

Крім того, все більше уваги приділяється інтеграції різних джерел даних, таких як погодні умови, економічні показники та демографічні зміни, що впливають на попит на енергію. Такий підхід дозволяє створювати комплексні моделі, які враховують багатофакторні впливи на енергоспоживання.

У багатьох країнах розробляються національні програми та ініціативи з прогнозування енергоспоживання, які підтримуються державою та приватним сектором. Це сприяє створенню нових інструментів та платформ для ефективного планування енергетичних ресурсів. Таким чином, наукові дослідження в галузі прогнозування енергоспоживання активно розвиваються, використовуючи передові технології та підходи для забезпечення точних та надійних прогнозів, що є основою для прийняття ефективних управлінських рішень у сфері енергетики.

Дослідженнями в області енергоспоживання займалися також в Тернопільському національному технічному університеті імені Івана Пулюя. Зокрема в роботі [1] описано створення програмних засобів моніторингу та прогнозування енергоспоживання, в роботі [2] і [3] вирішують задачу моделювання енергоспоживання, в [4], [5] та [6] вирішують аналогічну задачу

для споживання газу, роботи [7], [8], [9], [10] присвячені процесам енергозбереження в різних умовах.

Машинне навчання відіграє ключову роль у сучасних технологіях прогнозування енергоспоживання, забезпечуючи значний приріст точності та ефективності порівняно з традиційними методами прогнозування. Наведемо перелік деяких аспектів, котрі підкреслюють цю роль:

1. Аналіз великих обсягів даних: машинне навчання дозволяє аналізувати великі масиви даних, що включають історичні дані про споживання енергії, погодні умови, економічні індикатори та інші важливі фактори. Це дозволяє побудувати складні моделі, які враховують багатофакторні залежності.

2. Прогнозування на основі патернів: алгоритми машинного навчання можуть виявляти та використовувати складні патерни у споживанні енергії, які можуть бути невидимими для традиційних статистичних методів. Наприклад, нейронні мережі можуть самостійно виявляти та використовувати неочевидні взаємозв'язки між різними змінними.

3. Адаптивність до змін: моделі машинного навчання можуть бути легко адаптовані до мінливих умов, таких як зміна клімату або економічні коливання. Вони можуть автоматично підлаштовуватись до нових даних, що робить їх більш гнучкими і надійними для прогнозування на реальних часових шкалах.

4. Енергоефективність: застосування точних прогнозів енергоспоживання дозволяє ефективніше управляти виробництвом енергії та розподілом ресурсів. Це сприяє зменшенню втрат і оптимізації використання енергетичних ресурсів, що має велике значення для сталого розвитку.

Отже, машинне навчання відіграє важливу роль у вдосконаленні прогнозування енергоспоживання, що дозволяє ефективніше управляти енергетичними ресурсами та підтримувати стаке функціонування енергетичних систем.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Аналіз глобального попиту на електроенергію

Падіння споживання електроенергії в розвинутих економіках стримувало зростання світового попиту на електроенергію в 2023 році. Світовий попит на електроенергію зріс на 2,2% у 2023 році, що менше, ніж зростання на 2,4%, яке спостерігалось в 2022 році [11]. У той час як Китай, Індія та численні країни Південно-Східної Азії зазнали значного зростання. Що стосується попиту на електроенергію в 2023 році, країни з розвинутою економікою відзначили значне зниження через невтішне макроекономічне середовище та високу інфляцію, що призвело до скорочення виробництва та промислового виробництва.

Очікується, що глобальний попит на електроенергію зростатиме швидшими темпами протягом наступних трьох років, зростаючи в середньому на 3,4% щорічно до 2026 року. Досягнення буде зумовлено покращенням економічних прогнозів, що сприятиме швидшому зростанню попиту на електроенергію як у розвинутому і країни з економікою, що розвивається. Зокрема, у країнах з розвинутою економікою та Китаї попит на електроенергію підтримуватиметься триваючою електрифікацією житлового та транспортного секторів, а також значним розширенням сектора центрів обробки даних. За оцінками, частка електроенергії в кінцевому споживанні енергії досягла 20% у 2023 році проти 18% у 2015 році. Незважаючи на те, що це прогрес, електрифікацію потрібно швидко прискорити, щоб досягти світових цілей декарбонізації. У сценарії чистих нульових викидів до 2050 року МЕА, спрямованому на обмеження глобального потепління до 1,5 °C, частка електроенергії в кінцевому споживанні енергії наближається до 30% у 2030 році [11].

Споживання електроенергії центрами обробки даних, штучним інтелектом (ШІ) і сектором криптовалют може подвоїтися до 2026 року. Центри обробки даних є значними рушійними силами зростання попиту на електроенергію в

багатьох регіонах. Після глобального споживання 460 терават-годин (ТВт-год) у 2022 році загальне споживання електроенергії центрами обробки даних може досягти понад 1000 ТВт-год у 2026 році. Цей попит приблизно еквівалентний споживанню електроенергії в Японії. Оновлені правила та технологічні вдосконалення, включно з ефективністю, матимуть вирішальне значення для пом'якшення різкого зростання споживання енергії центрами обробки даних.

1.2 Країни, що розвиваються, є двигунами зростання глобального попиту на електроенергію

Близько 85% додаткового попиту на електроенергію до 2026 року буде надходити з-за меж розвинутих економік, причому Китай зробить значний внесок, незважаючи на те, що економіка країни зазнає структурних змін. У 2023 році попит на електроенергію в Китаї зріс на 6,4%, завдяки секторам послуг і промисловості. Оскільки економічне зростання країни, як очікується, сповільниться і стане менш залежним від важкої промисловості, за нашими прогнозами, темпи зростання попиту на електроенергію в Китаї зменшаться до 5,1% у 2024 році, 4,9% у 2025 році та 4,7% у 2026 році [11]. Незважаючи на це, загальне збільшення попиту на електроенергію в Китаї до 2026 року приблизно на 1400 ТВт-год становить більше половини поточного річного споживання електроенергії в Європейському Союзі. Споживання електроенергії на душу населення в Китаї наприкінці 2022 року вже перевищило показники Європейського Союзу та має намір зростати й далі. Швидке розширення виробництва сонячних фотомодулів і електромобілів, а також обробка супутніх матеріалів сприятиме постійному зростанню попиту на електроенергію в Китаї в той час як структура його економіки розвиватиметься.

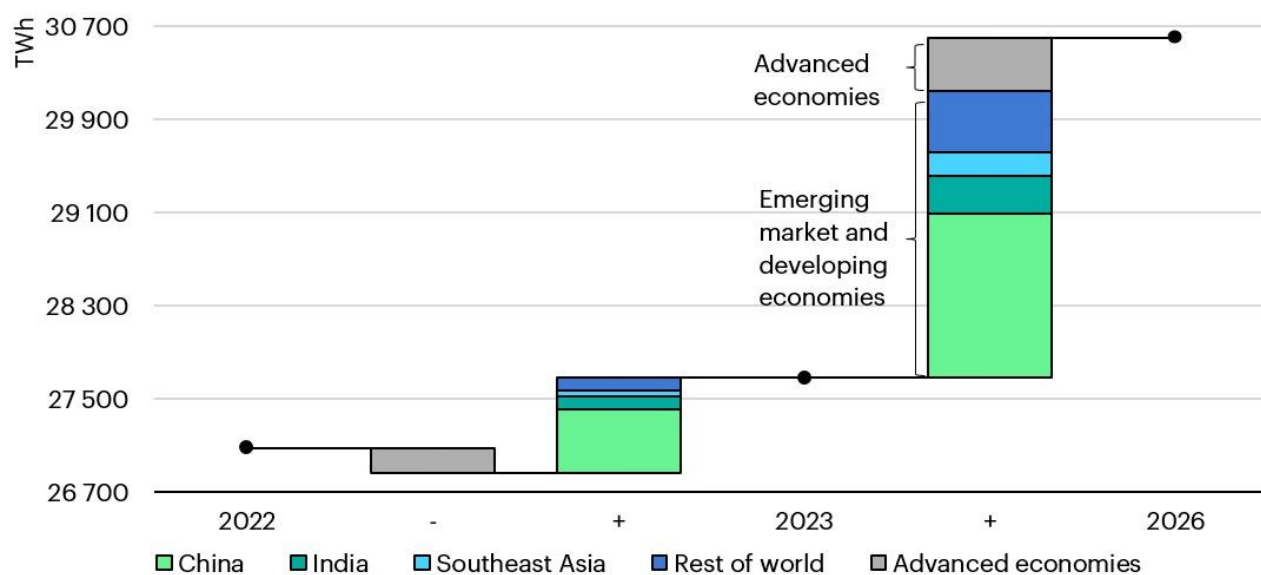


Рисунок 1.1 – Зміна попиту на електроенергію по регіонах, 2022-2026 рр

Китай забезпечує найбільшу частку світового зростання попиту на електроенергію з точки зору обсягу, але Індія демонструє найшвидші темпи зростання до 2026 року серед великих економік. Після збільшення попиту на електроенергію в Індії на 7% у 2023 році ми очікуємо зростання понад 6% щорічно до 2026 року за підтримки високої економічної активності та розширення власності на кондиціонери. Протягом наступних трьох років Індія збільшить попит на електроенергію, приблизно еквівалентний поточному споживанню Сполученого Королівства. Хоча відновлювані джерела енергії мають задовольнити майже половину цього зростання попиту, очікується, що одна третина надходитиме від зростання виробництва електроенергії на вугіллі. Ми також очікуємо, що до 2026 року в Південно-Східній Азії спостерігатиметься значне щорічне зростання попиту на електроенергію в середньому на 5%, що буде зумовлено високою економічною активністю.

У той час як споживання електроенергії на душу населення в Індії та Південно-Східній Азії стрімко зростає, в Африці воно фактично стагнує вже більше трьох десятиліть. Споживання на душу населення в Африці навіть знизилося в останні роки, оскільки населення зростало швидше, ніж стало доступним постачання електроенергії, і ми очікуємо, що воно відновиться до рівня 2010-15 років не раніше кінця 2026 року. Тридцять років тому житель

Африки споживав у середньому більше електроенергії, ніж житель Індії чи Південно-Східної Азії. Однак значне зростання попиту та пропозиції електроенергії в Індії та Південно-Східній Азії за останні десятиліття, яке йшло рука об руку з бумом економічного розвитку, трансформувало ці регіони вражаючими темпами. Тим часом споживання електроенергії на душу населення в Африці в 2023 році було вдвічі менше, ніж в Індії, і на 70% нижче, ніж у Південно-Східній Азії. Наш прогноз для Африки на період 2024-2026 рр. передбачає середньорічне зростання загального попиту на електроенергію на 4%, що вдвічі перевищує середній темп зростання, який спостерігався між 2017 і 2023 роками. Дві третини цього зростання попиту буде задоволено за рахунок розширення відновлюваних джерел енергії, а решта покривається переважно природним газом.

Попит на електроенергію в Сполучених Штатах впав на 1,6% у 2023 році після зростання на 2,6% у 2022 році, але очікується, що він відновиться в перспективному періоді 2024-26. Основною причиною падіння стала більш м'яка погода у 2023 році порівняно з 2022 роком, хоча фактором також було уповільнення у виробничому секторі. Ми прогнозуємо помірне зростання попиту на 2,5% у 2024 році, припускаючи повернення до середніх погодних умов. За цим послідує зростання в середньому на 1% у 2025-26 роках, головним чином за рахунок електрифікації та розширення сектору центрів обробки даних, на який, як очікується, припадатиме більше однієї третини додаткового попиту до 2026 року.

1.3 Перспективи на швидке відновлення енергоємних галузей у Європейському Союзі

У 2023 році попит на електроенергію в Європейському Союзі впав другий рік поспіль, незважаючи на те, що ціни на енергоносії впали з рекордно високого рівня. Після падіння на 3,1% у 2022 році, зниження попиту в ЄС на 3,2% у річному вимірі у 2023 році означало, що він знизився до рівня, який востаннє спостерігався два десятиліття тому. Як і в 2022 році, нижче споживання в

промислового секторі стало основним фактором, який знизив попит на електроенергію, оскільки ціни на енергію знизилися, але залишилися вищими за рівень до пандемії. У 2023 році також спостерігалися ознаки деякого постійного зниження попиту, особливо в енергоємній хімічній галузі та виробництві первинних металів. Ці сегменти залишаються вразливими до цінових шоків на енергоносії протягом нашого прогнозованого періоду.

Очікується, що споживання електроенергії в ЄС повернеться до рівня 2021 року не раніше 2026 року. Попит на електроенергію в промисловому секторі Європейського Союзу впав приблизно на 6% у 2023 році після аналогічного падіння у 2022 році. Якщо припустити, що промисловий сектор поступово відновиться, оскільки ціни на енергоносії будуть помірними, прогнозується, що зростання попиту на електроенергію в ЄС зросте в середньому на 2,3% у 2024-2026 роках. Електричні транспортні засоби, теплові насоси та центри обробки даних залишаються міцними стовпами зростання протягом цього періоду – разом припадаючи на половину очікуваного зростання загального попиту.

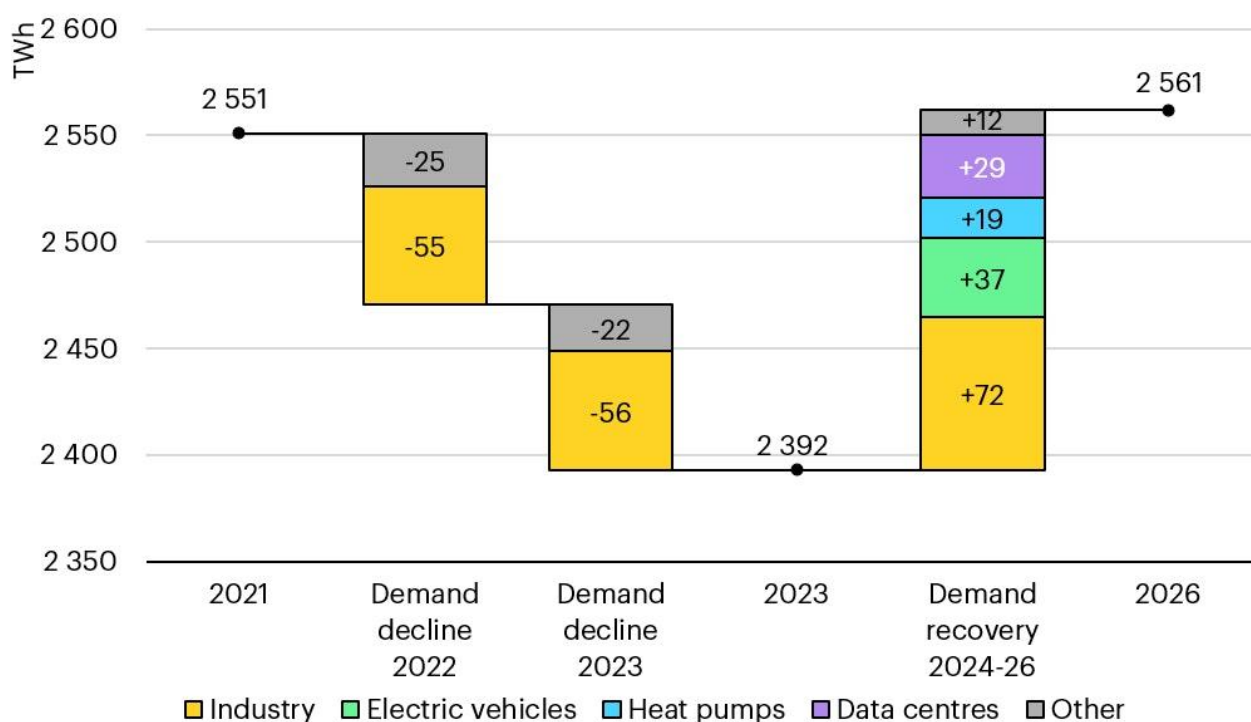


Рисунок 1.2 – Очікувані чинники зміни попиту на електроенергію в Європейському Союзі, 2021-2026 рр.

Ціни на електроенергію для енергоємних виробництв у Європейському Союзі у 2023 році були майже вдвічі вищими, ніж у США та Китаї. Незважаючи на зниження цін у Європейському Союзі на 50% у 2023 році порівняно з 2022 роком, після вторгнення Росії в Україну енергоємні галузі в регіоні продовжували стикатися зі значно вищими витратами на електроенергію порівняно зі Сполученими Штатами та Китаєм. Ціновий розрив між енергоємними галузями в Європейському Союзі та США та Китаї, який існував до енергетичної кризи, збільшився. У результаті очікується, що конкурентоспроможність енергоємних галузей ЄС залишатиметься під тиском. Політики зараз обговорюють нові політичні ініціативи та фінансові інструменти, які дозволять Європейському Союзу позиціонувати себе серед інших глобальних промислових важковаговиків. Масштаби та ефективність цих заходів, ймовірно, визначатимуть майбутнє енергоємного промислового сектора Європейського Союзу.

1.4 Прогноз постачання чистої електроенергії

Рекордне виробництво електроенергії з джерел з низьким рівнем викидів, включаючи атомну та відновлювану енергію, як-от сонячна, вітрова та гідроелектростанція, має покрити весь глобальний попит протягом наступних трьох років. За прогнозами, до 2026 року на джерела з низьким рівнем викидів, які зменшать роль викопного палива у виробництві електроенергії в усьому світі, припадатиме майже половина світового виробництва електроенергії, порівняно з 39% у 2023 році. Протягом наступних трьох років у період з 2018 по 2023 рік виробництво з низьким рівнем викидів зросте вдвічі вище, ніж щорічні темпи зростання – наслідок зміни, враховуючи, що енергетичний сектор робить найбільший внесок у глобальний вуглекислий газ (CO₂). викидів сьогодні.

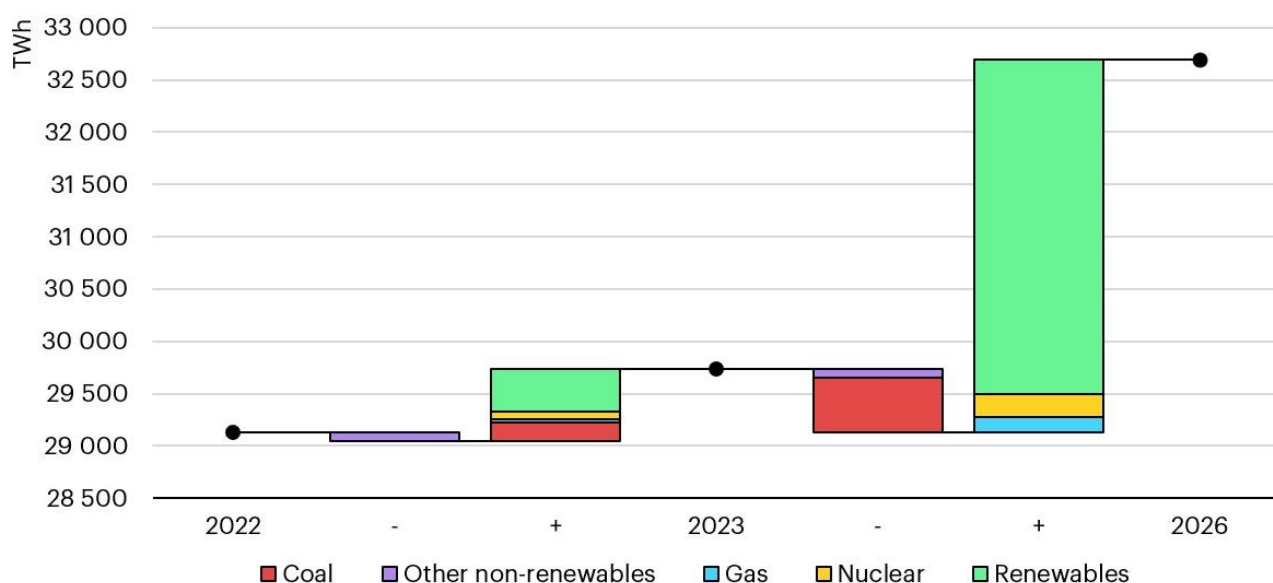


Рисунок 1.3 – Зміни у світовому виробництві електроенергії, 2022-2026

До початку 2025 року відновлювані джерела енергії забезпечуватимуть більше третини загального виробництва електроенергії у світі, обігнавши вугілля. Прогнозується, що частка відновлюваних джерел енергії у виробництві електроенергії зросте з 30% у 2023 році до 37% у 2026 році, при цьому зростання значною мірою буде підтримуватися розширенням дедалі дешевшої сонячної фотоелектричної енергії. Протягом цього періоду відновлювані джерела енергії з лишком компенсуватимуть зростання попиту в країнах із розвинутою економікою, таких як Сполучені Штати та Європейський Союз, витісняючи пропозицію з використанням викопного палива. У той же час очікується, що в Китаї швидке розширення відновлюваних джерел енергії задовольнить увесь додатковий попит на електроенергію, хоча погодні умови та ступінь уповільнення зростання попиту в країні залишаються ключовими джерелами невизначеності для прогнозів. Значне розширення потужностей відновлюваної електроенергії також має супроводжуватися прискореними інвестиціями в мережі та гнучкість системи, щоб забезпечити її плавну інтеграцію.

Швидке зростання відновлюваних джерел енергії, яке підтримується зростанням ядерної генерації, має витіснити світову вугільну генерацію, яка, за прогнозами, скорочуватиметься в середньому на 1,7% щорічно до 2026 року. Це слідє за збільшенням виробництва вугілля на 1,6% у 2023 році. на тлі посух в

Індії та Китаї, які зменшили виробництво гідроелектроенергії та збільшили виробництво електроенергії на вугіллі, з лишком компенсуючи значне зниження виробництва електроенергії на вугіллі в Сполучених Штатах та Європейському Союзі. Головним чинником, який визначатиме глобальні перспективи, є тенденції, що розвиваються в Китаї, де відбувається більше половини світового виробництва вугільної енергії. Вугільна генерація в Китаї наразі переживає повільний структурний спад, спричинений потужним розширенням відновлюваних джерел енергії та зростанням ядерної генерації, а також уповільненням економічного зростання. Незважаючи на введення в експлуатацію нових станцій для підвищення безпеки енергопостачання, очікується, що рівень використання китайських вугільних електростанцій продовжить падати, оскільки вони використовуються більш гнучко для доповнення до відновлюваних джерел. Тим не менш, на вугільну генерацію в Китаї значною мірою впливатимуть темпи ребалансування економіки, тенденції розвитку гідроенергетики та вузькі місця в інтеграції відновлюваних джерел енергії в енергетичну систему країни.

Очікується, що виробництво електроенергії на природному газі дещо зросте протягом прогнозного періоду. У 2023 році різке скорочення виробництва електроенергії на газі в Європейському Союзі було з лишком компенсовано величезним зростанням у Сполучених Штатах, де природний газ, який дедалі більше замінює вугілля, зафіксував найвищу частку виробництва електроенергії за всю історію. У 2023 році світове виробництво газу зросло менше ніж на 1%. До 2026 року ми прогнозуємо середньорічний темп зростання приблизно на 1%. У той час як очікується, що виробництво газу в Європі продовжить скорочуватися, глобальне зростання буде підтримуватися значним зростанням в Азії, на Близькому Сході та в Африці на тлі зростання попиту на електроенергію в цих регіонах і наявності додаткових поставок зрідженого природного газу (СПГ) з 2025 і далі.

1.5 Перспективи виробництва ядерної енергії

Прогнозується, що до 2025 року світова ядерна генерація перевищить попередній рекорд, встановлений у 2021 році. Незважаючи на те, що деякі країни поступово припиняють роботу атомної електростанції або достроково виводять із експлуатації електростанції, прогнозується, що до 2026 року виробництво атомної енергії зростатиме в середньому на 3% на рік, оскільки у Франції завершаться роботи з технічного обслуговування, Японія відновить ядерне виробництво на кількох електростанціях і побудує нові реактори. Почати комерційну діяльність на різних ринках, включаючи Китай, Індію, Корею та Європу. Багато країн роблять атомну енергетику важливою частиною своїх енергетичних стратегій, оскільки вони прагнуть забезпечити енергетичну безпеку, одночасно зменшуючи викиди парникових газів.

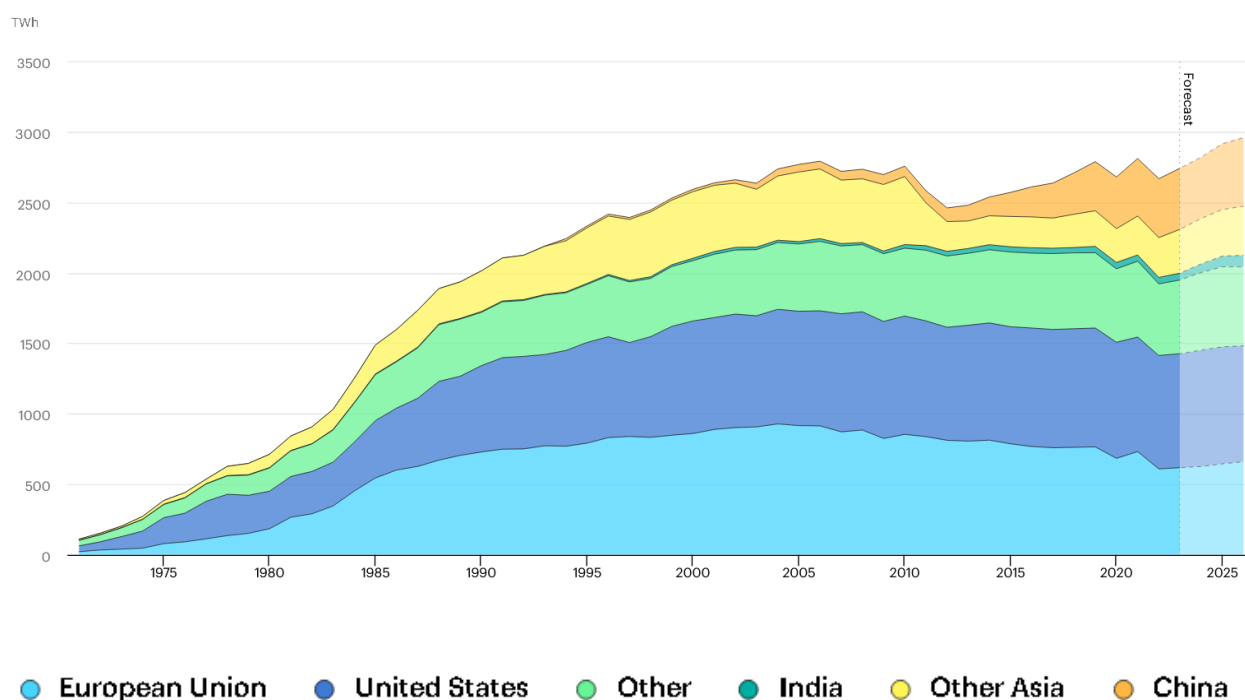


Рисунок 1.4 – Еволюція виробництва атомної енергії за регіонами, 1972-2026

На конференції зі зміни клімату COP28, яка завершилася в грудні 2023 року, понад 20 країн підписали спільну декларацію про потроєння потужностей ядерної енергетики до 2050 року. Досягнення цієї мети вимагатиме вирішення основного завдання – зниження будівельних і фінансових ризиків у ядерному

секторі. Імпульс також зростає за технологією малих модульних реакторів (SMR). Розвиток і розгортання технології залишається скромним і не позбавленим труднощів, але дослідження та розробки починають набирати обертів.

Азія залишається основним рушієм зростання ядерної енергетики, з прогнозується, що частка регіону в світовому виробництві атомної енергії досягне 30% у 2026 році. До кінця 2026 року Азія має випередити Північну Америку як регіон з найбільшою встановленою ядерною потужністю, причому очікується завершення будівництва великої кількості станцій, які зараз будуються до того часу. Більше половини нових реакторів, які, як очікується, запрацюють протягом прогнозованого періоду, знаходяться в Китаї та Індії. За останнє десятиліття ядерна енергетика в Китаї особливо сильно зросла, збільшивши потужність приблизно на 37 гігават (ГВт), що еквівалентно майже двом третинам поточної ядерної потужності. Це призвело до того, що частка Китаю у світовій ядерній генерації зросла з 5% у 2014 році до приблизно 16% у 2023 році. У грудні 2023 року Китай розпочав комерційну експлуатацію свого першого реактора четвертого покоління, що ще більше підкреслило розвиток атомної енергетики країни.

Викиди від виробництва електроенергії вступають у структурний занепад, оскільки декарбонізація набирає темпи

Глобальні викиди CO₂ від виробництва електроенергії впадуть більш ніж на 2% у 2024 році після збільшення на 1% у 2023 році [11]. За цим планується невелике зниження у 2025 та 2026 роках. Значне зростання виробництва електроенергії на вугіллі у 2023 році – особливо в Китаї та Індії на тлі скорочення виробництва гідроенергії – спричинило зростання викидів CO₂ у світовому секторі електроенергетики. Оскільки постачання чистої електроенергії продовжує швидко зростати, прогнозується, що частка викопного палива у світовій генерації знизиться з 61% у 2023 році до 54% у 2026 році, вперше впавши нижче 60% за даними МЕА з 1971 року. погодні умови, економічні потрясіння або зміни в урядовій політиці можуть призвести до тимчасового зростання

викидів в окремі роки, очікується, що більш широке зниження викидів у енергетичному секторі продовжуватиметься, оскільки відновлювані джерела енергії та потужності атомної енергетики продовжуватимуть розширюватися та витісняти виробництво, що працює на викопному паливі (рис. 1.5 – 1.8).

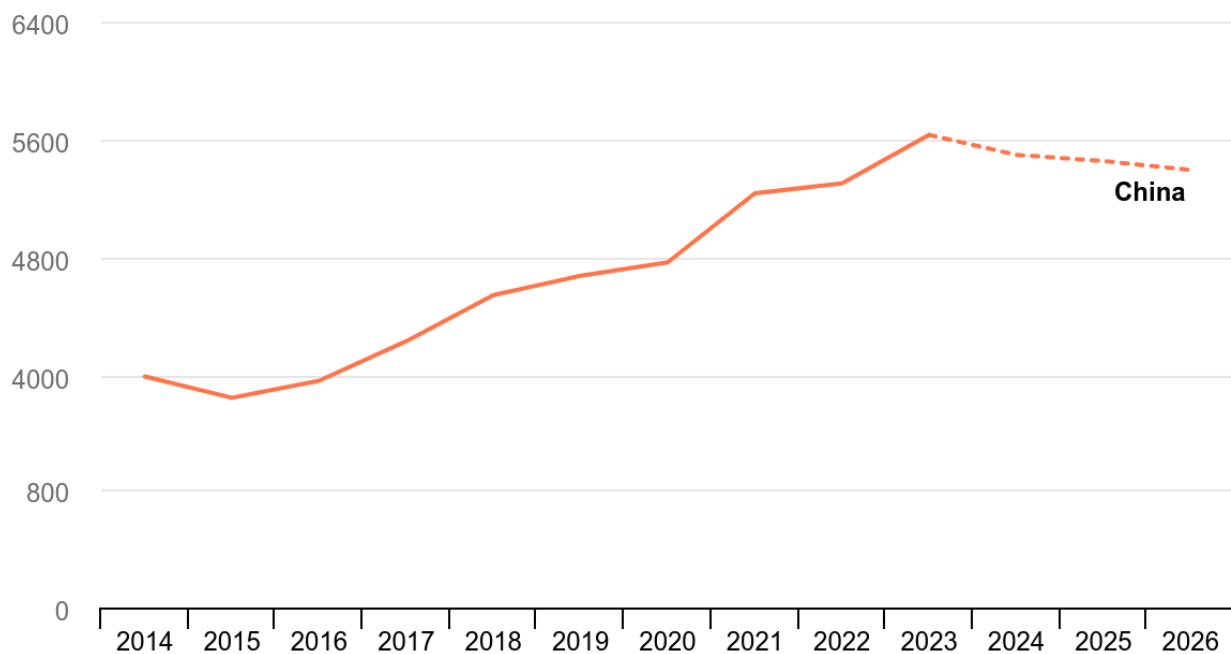


Рисунок 1.5 – Викиди CO₂ від виробництва електроенергії в Китаї, 2014-2026

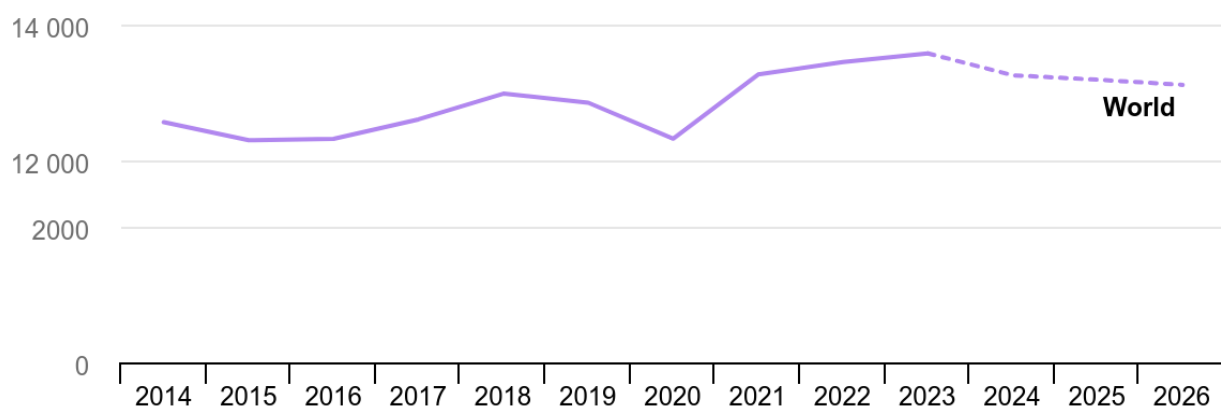


Рисунок 1.6 – Глобальні викиди CO₂ від виробництва електроенергії, 2014-2026 рр.

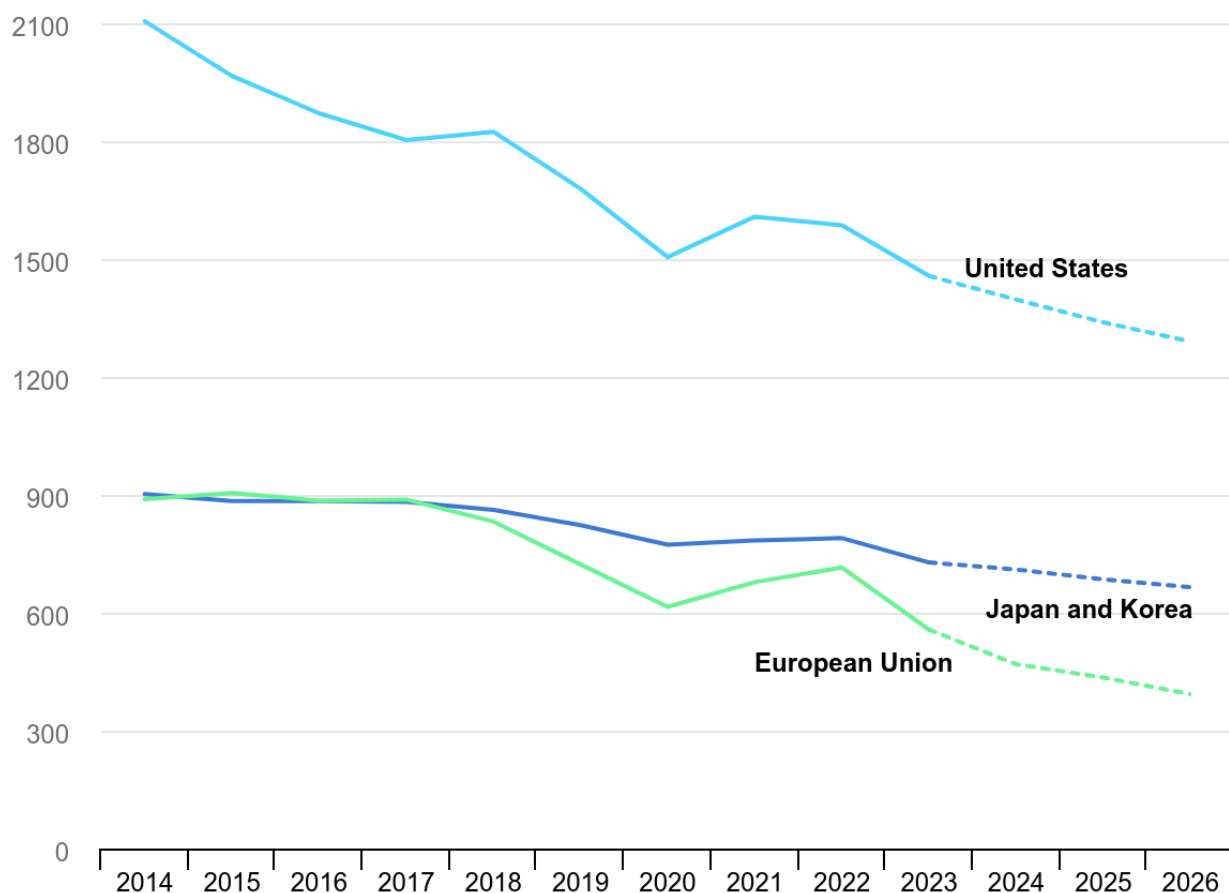


Рисунок 1.7 – Викиди CO₂ від виробництва електроенергії в окремих регіонах, 2014 - 2026 рр.

CO₂ у світовому виробництві електроенергії скоротиться вдвічі, ніж у період до пандемії. Прогнозований середній спад на 4% в інтенсивності викидів CO₂ між 2023 і 2026 роками вдвічі перевищує 2%, що спостерігався в період між 2015 і 2019 роками. Очікується, що Європейський Союз зафіксує найвищий прогрес у скороченні інтенсивності викидів, у середньому покращивши 13% на рік. Далі йдуть Китай з прогнозованим щорічним покращенням на 6% і Сполучені Штати на 5%. Зменшення інтенсивності викидів CO₂ при виробництві електроенергії означає, що економія викидів за рахунок електрифікації транспорту, опалення та промисловості стане ще більшою.

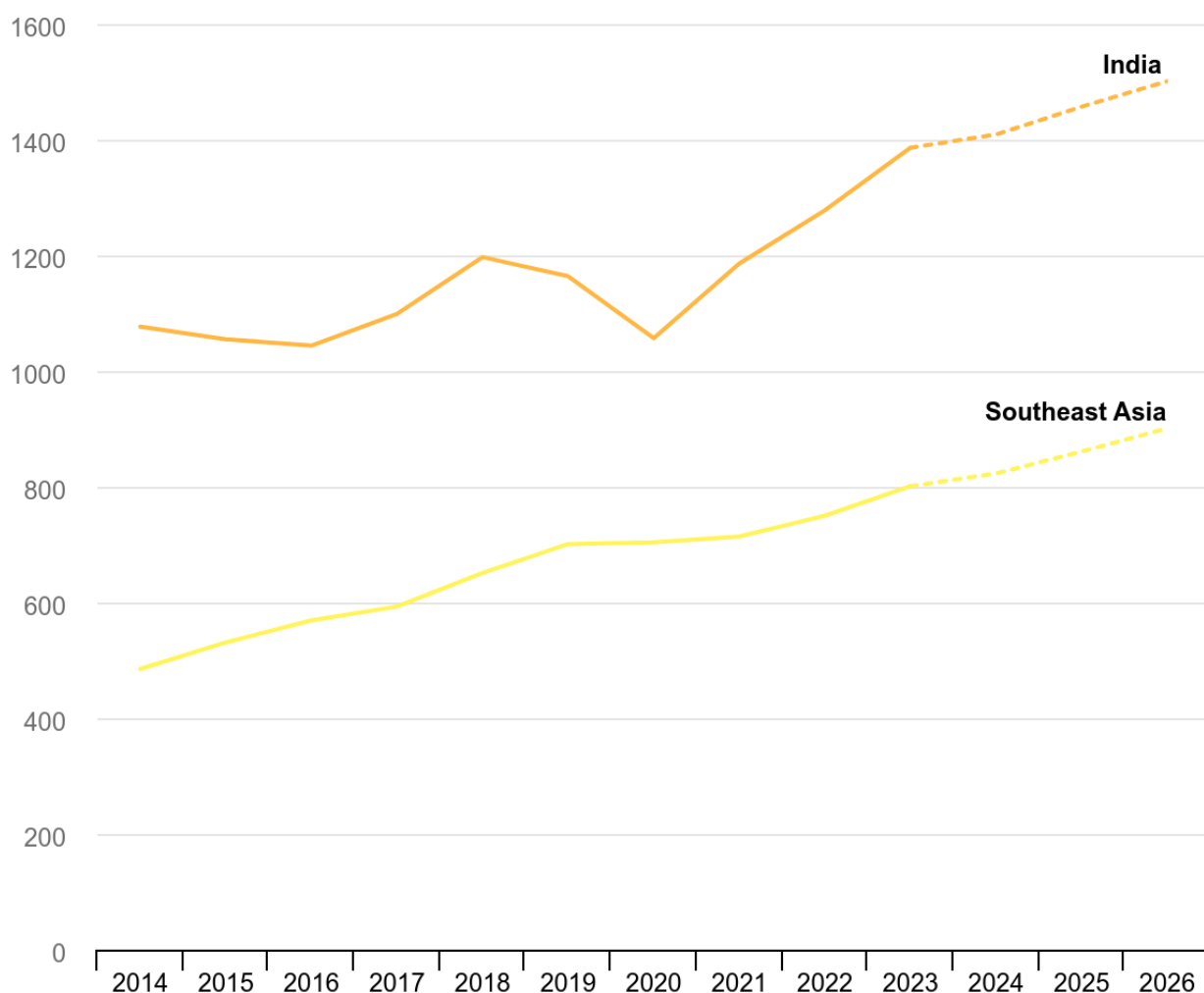


Рисунок 1.8 – Викиди CO₂ від виробництва електроенергії в Індії та Південно-Східній Азії, 2014-2026 рр.

Оптові ціни на електроенергію в багатьох країнах у 2023 році впали з рекордно високих значень, зафіксованих у 2022 році. Це відбулося одночасно зі зниженням цін на енергетичні товари, такі як природний газ і вугілля. Однак існують регіональні відмінності. Оптові ціни на електроенергію в Європі в 2023 році впали в середньому більш ніж на 50% порівняно з рекордними рівнями 2022 року. Незважаючи на це, ціни в Європі все ще були приблизно вдвічі вищими за рівні 2019 року, тоді як ціни в США в 2023 році були лише приблизно на 15% вищими, ніж у 2019 році. Невизначеність про темпи відновлення ядерної енергетики Франції та ціни на природний газ сприяють зростанню ф'ючерсних

цін у Європі на наступні зими. Країни Північної Європи, де домінує гідроенергетика, залишаються єдиним ринком в Європі, де середні оптові ціни на електроенергію можна порівняти з цінами в Сполучених Штатах і Австралії. У 2023 році оптові ціни в Японії та Індії також залишалися вищими за рівень 2019 року.

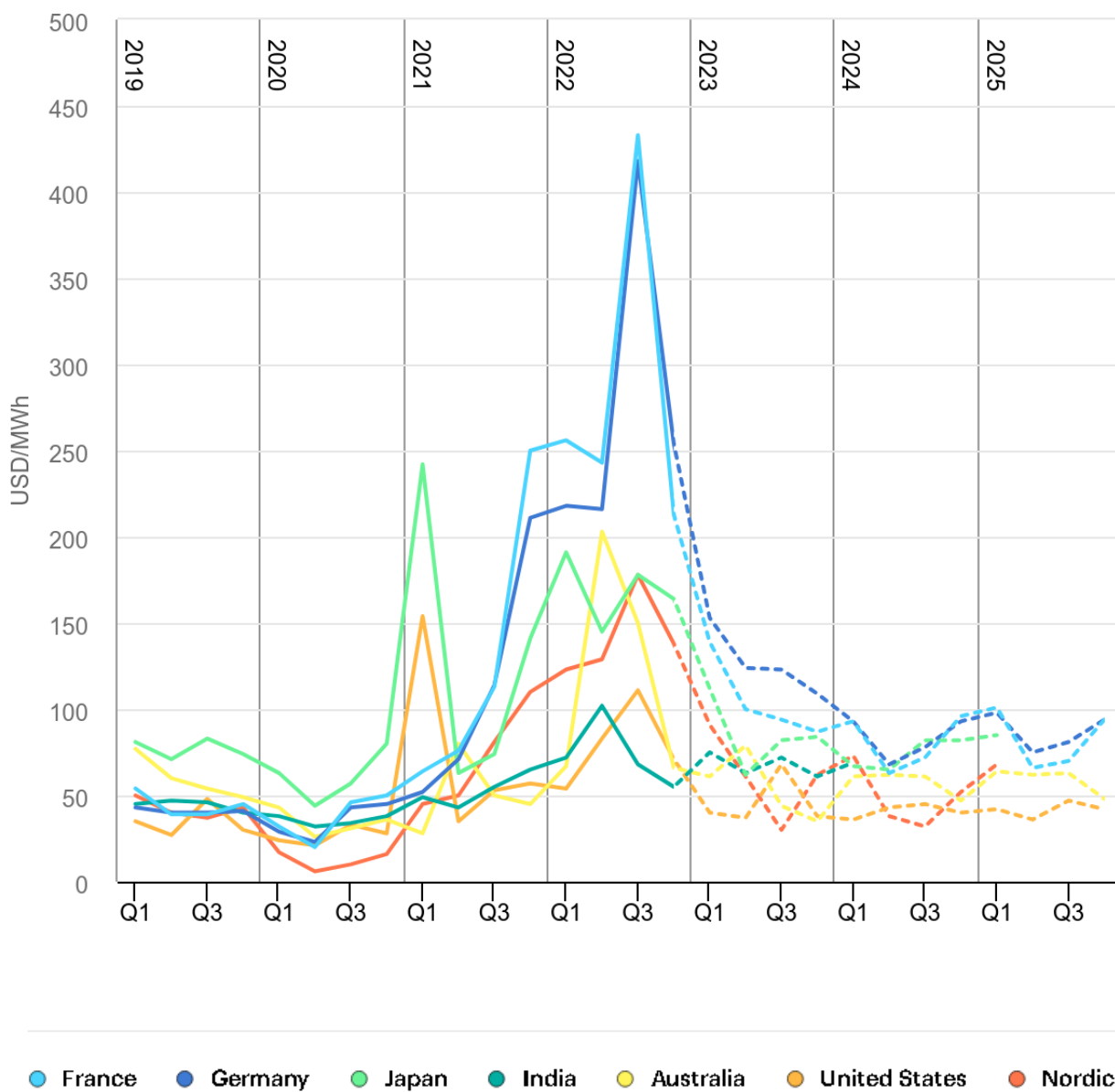


Рисунок 1.9 – Середньоквартальні оптові ціни для окремих регіонів, 2019-2025 рр

Зростаючий вплив погоди на енергосистеми підкреслює важливість інвестицій у безпеку електроенергії

Глобальне виробництво гідроелектроенергії скоротилося в 2023 році через погодні умови, такі як посухи, кількість опадів нижче середньої та раннє танення

снігу в багатьох регіонах. У Канаді, Китаї, Колумбії, Коста-Ріці, Індії, Мексиці, Туреччині, Сполучених Штатах і В'єтнамі, разом з іншими країнами, спостерігалось зниження виробництва гідроенергії. Коефіцієнт глобальної гідроенергетичної потужності, ключовий показник рівня використання, впав нижче 40%, що є найнижчим значенням, зареєстрованим щонайменше за три десятиліття. У деяких країнах зменшення виробництва гідроелектроенергії призвело до дефіциту енергії, збільшення залежності від викопних джерел, таких як вугілля та газ, і викликало занепокоєння щодо стабільності постачання електроенергії. Загальна тенденція підкреслює сприйнятливість гідроенергетики до погодних умов і потенційного впливу країн, які значною мірою покладаються на гідроенергію для виробництва електроенергії. Диверсифікація джерел енергії, побудова регіональних енергетичних зв'язків і впровадження стратегій стійкої генерації в умовах зміни погодних умов ставатимуть все більш важливими.

У 2023 році в Сполучених Штатах та Індії екстремальні погодні явища спричинили серйозні відключення електроенергії. Це підкреслило необхідність підвищити стійкість, оскільки вплив погодних умов на енергетичні системи зростає, а попит і пропозиція стають все більш залежними від погоди. Недостатня потужність електроенергії, проблеми з постачанням палива та технічні проблеми, пов'язані з мережею, також продовжували спричиняти значні дефіцити електроенергії в багатьох регіонах. Більшість цих збоїв спостерігалась в країнах з економікою, що розвивається, наприклад Пакистані, Кенії та Нігерії, які особливо страждають від недостатнього постачання електроенергії, проблем з інфраструктурою та напружених мереж в умовах зростання попиту на електроенергію. Розширені, потужніші мережі не тільки забезпечуватимуть надійну електроенергію, але й стануть життєво важливою опорою для інтеграції відновлюваних джерел енергії в енергетичні системи. Удосконалення збору даних, оцифрування та більша прозорість даних щодо збоїв також є важливими для забезпечення кращого розуміння причин виникнення несправностей і для розробки запобіжних заходів.

Спеціальні операційні заходи та нові ринки для забезпечення стабільності енергосистем стають все більш поширеними. Країни з високою часткою змінної відновлюваної генерації впроваджують механізми для забезпечення сталої частоти енергосистеми. Деякі регіони встановлюють мінімальні вимоги до інерції системи, властивості, яка зазвичай забезпечується звичайними генераторами з обертовими роторами, що допомагає підвищити стійкість енергосистеми під час збоїв. Крім того, різні країни, включаючи Сполучене Королівство, Ірландію та Австралію, запроваджують ринки та заходи, такі як швидка частотна характеристика та подібні послуги, які швидко стабілізують енергосистему після збоїв. Системи зберігання акумуляторів можуть надавати такі послуги для стабільності мережі, одночасно підвищуючи гнучкість системи, таким чином відіграючи вирішальну роль в інтеграції відновлюваних джерел енергії.

Отже прогнози енергоспоживання – важлива задача. В даній роботі вирішимо її для масштабів невеликого населеного пункту.

2 ОГЛЯД АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ЕНЕРГОСПОЖИВАННЯ

2.1 Принципи використання AutoML

AutoML кардинально змінює правила машинного навчання. Це навчання моделей машинного навчання для автоматизації процесу вибору та налаштування алгоритмів. Це включає все, починаючи від попередньої обробки даних і закінчуючи вибором найбільш підходящої моделі для поставленого завдання. Інструменти AutoML вирішують завдання налаштування гіперпараметрів і вибору моделі, які зазвичай потребують часу та досвіду. За допомогою AutoML користувачі без досвіду машинного навчання можуть тренувати високоефективні моделі з мінімальними зусиллями. Незалежно від того, чи є ви власником малого бізнесу, дослідником чи фахівцем із обробки даних, AutoML допоможе досягти цілей з меншими витратами часу та зусиль. Приклади популярних платформ AutoML включають Google Cloud AutoML, H2O.ai і DataRobot.

AutoML забезпечує зрозумілий ШІ для покращення інтерпретації моделі. Це дозволяє дослідникам даних зрозуміти, як модель робить прогнози, що особливо корисно в охороні здоров'я, фінансах та автономних системах. Це можна використовувати для виявлення упередженості в даних і запобігання неправильним прогнозам. Наприклад, AutoML можна використовувати в охороні здоров'я для діагностики захворювань шляхом аналізу медичних зображень, у фінансовій сфері для виявлення шахрайства, у роздрібній торгівлі для рекомендацій продуктів і в транспорті для безпілотних автомобілів. На рисунку 2.1 показано процес AutoML.

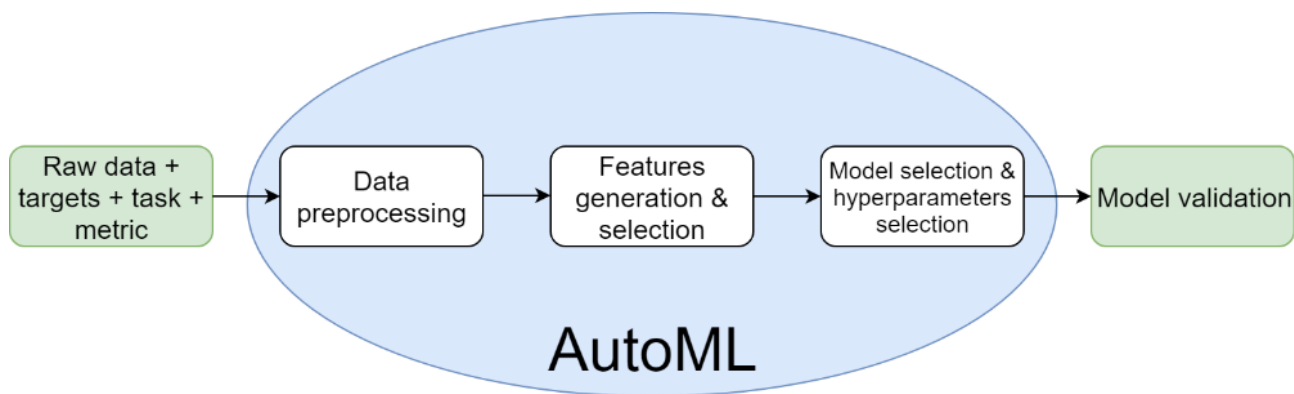


Рисунок 2.1 – Процес AutoML

AutoML автоматизує використання машинного навчання для реальних проблем. Це включає такі завдання, як вибір алгоритму, оптимізація гіперпараметрів і розробка функцій.

2.1.1 Огляд AutoML

Для вирішення різних аспектів проблеми розробляється багато різних методів. Деякі популярні підходи наведено нижче:

- Пошук нейронної архітектури (NAS): цей метод використовує алгоритм пошуку для автоматичного пошуку найкращої архітектури нейронної мережі для заданого завдання та набору даних.
- Байєсова оптимізація: цей метод використовує імовірнісну модель для керування пошуком найкращого набору гіперпараметрів для заданої моделі та набору даних.
- Еволюційні алгоритми: цей метод використовує еволюційні алгоритми, такі як генетичні алгоритми або оптимізація рою частинок, для пошуку найкращого набору гіперпараметрів моделі.
- Методи на основі градієнта: цей метод використовує методи оптимізації на основі градієнта, такі як градієнтний спуск, Адам тощо, для оптимізації гіперпараметрів моделі.
- Передача навчання: цей метод використовує попередньо навчену модель для подібного завдання або набору даних як початкову точку, а потім налаштовує її на цільове завдання та набір даних.

- **Методи ансамблю:** цей метод поєднує кілька моделей для створення більш надійної та точної кінцевої моделі.
- **Багатомодальні методи:** цей метод використовує кілька модальностей даних, таких як зображення, текст і аудіо, для навчання моделей і підвищення продуктивності.
- **Метанавчання:** цей метод використовує модель, щоб дізнатися, як вчитися на даних, що може підвищити ефективність процесу вибору моделі.
- **Одноразове або кількакратне навчання:** за допомогою цього методу можна навчитися розпізнавати нові класи лише на одному чи кількох прикладах.

На ринку доступно багато різних інструментів і бібліотек AutoML, таких як AutoML від Google, AutoML від H2O.ai, DataRobot, TPOT і auto-sklearn. Ці інструменти надають широкий спектр функцій, які можна інтегрувати в існуючі робочі процеси.

2.1.2 Переваги AutoML у машинному навчанні

AutoML спрощує процес машинного навчання та забезпечує багато переваг, деякі з яких наведено нижче.

- **Економія часу.** Автоматизація процесу вибору моделі та налаштування гіперпараметрів може заощадити значну кількість часу для спеціалістів із обробки даних та інженерів машинного навчання.
- **Доступність:** AutoML дозволяє користувачам з невеликим досвідом або зовсім без досвіду машинного навчання навчати високоефективні моделі.
- **Покращена продуктивність:** Методи AutoML часто можуть знайти кращі архітектури моделі та параметри гіперпараметрів, ніж ручні методи, що призводить до покращення продуктивності моделі.
- **Обробка великих обсягів даних:** AutoML може обробляти великі обсяги даних і знаходити найкращу модель навіть із більшою кількістю функцій.
- **Масштабованість:** AutoML може масштабуватися до великих наборів даних і складних моделей, завдяки чому він добре підходить для великих даних і високопродуктивних обчислювальних середовищ.

- Універсальність: AutoML можна використовувати в різних галузях промисловості та застосуваннях, включаючи охорону здоров'я, фінанси, роздрібну торгівлю та транспорт.
- Економічно ефективним: AutoML може заощадити ресурси та гроші в довгостроковій перспективі, зменшивши потребу в ручній праці та досвіді.
- Зменшення ризику людської помилки: автоматизація вибору моделі та процесу налаштування гіперпараметрів може зменшити ризик людської помилки та покращити відтворюваність результатів.
- Підвищена ефективність: AutoML можна інтегрувати з іншими інструментами та процесами для підвищення ефективності конвеєра даних.
- Обробка кількох модальностей даних: AutoML може обробляти різні модальності даних, як-от зображення, текст і аудіо, для навчання моделей і підвищення продуктивності.

AutoML пропонує ряд переваг для спеціалістів із обробки даних та інженерів, які економлять час і ресурси завдяки автоматизації нудних і трудомістких завдань. Це також покращує інтерпретабельність моделі, надаючи зрозумілий ШІ. Ці сукупні переваги роблять AutoML цінним інструментом у багатьох галузях промисловості та застосуваннях.

2.1.3 Недоліки AutoML

AutoML став популярним інструментом для науковців і аналітиків даних. Однак він має обмеження. Існують наступні обмеження, наведені нижче.

- Обмежений контроль над процесом вибору моделі та налаштування гіперпараметрів: Методи AutoML працюють на основі попередньо визначених алгоритмів і налаштувань, і користувачі можуть мати обмежений контроль над остаточною моделлю.
- Обмежена інтерпретація отриманої моделі: Методи AutoML можуть бути непрозорими, що ускладнює розуміння того, як модель робить свої прогнози.

- Вищі витрати, ніж ручне проектування та навчання моделі: Інструменти та інфраструктура AutoML можуть бути дорогими для впровадження та обслуговування.

- Складність включення предметно-специфічних знань у модель: AutoML покладається на дані та попередньо визначені алгоритми, які можуть бути менш ефективними, якщо включати предметні знання.

- Потенційна низька продуктивність на граничних випадках або незвичний розподіл даних: Методи AutoML можуть погано працювати з даними, які значно відрізняються від навчальних даних.

- Обмежена підтримка для певних моделей або завдань: методи AutoML можуть не підходити для всіх моделей або завдань.

- Залежність від великої кількості позначених даних: Методи AutoML зазвичай вимагають великих обсягів позначених даних для ефективного навчання моделей.

- Обмежена можливість обробки даних із відсутніми значеннями або помилками: методи AutoML можуть погано працювати з даними з відсутніми значеннями чи помилками.

- Обмежена здатність пояснювати прогнози та рішення моделі: Методи AutoML можуть бути непрозорими, що ускладнює розуміння того, як модель робить свої прогнози, що може бути проблемою для певних програм і галузей.

- Переобладнання: Методи AutoML можуть призвести до переобладнання навчальних даних, якщо вони не контролюються належним чином, що може призвести до низької продуктивності нових невидимих даних.

AutoML – це потужний інструмент для автоматизації процесу машинного навчання, але він має свої обмеження. Важливо враховувати ці обмеження за наявності експертного нагляду для підтвердження результатів.

2.1.4 Практичне застосування автоматизації машинного навчання

Нижче наведено кілька практичних прикладів AutoML.

- AutoML Vision від Google дозволяє користувачам навчати власні моделі машинного навчання для розпізнавання зображень за допомогою своїх наборів даних зображень
- AutoML від H2O.ai дає змогу спеціалістам із обробки даних та аналітикам автоматично навчати та оптимізувати моделі машинного навчання без необхідності писати код
- DataRobot надає платформу AutoML, яка може автоматично створювати, оцінювати та розгортати моделі машинного навчання для широкого діапазону випадків використання, включаючи виявлення шахрайства, прогнозування відтоку клієнтів і прогнозне обслуговування
- Amazon SageMaker – це повністю керований сервіс, який дозволяє науковцям і розробникам даних швидко й легко створювати, навчати та розгортати моделі машинного навчання в масштабі
- IBM Watson AutoAI – це платформа, яка автоматизує процес побудови, навчання та розгортання моделей машинного навчання та надає функції інтерпретації та пояснення, які допомагають користувачам зрозуміти процеси прийняття рішень у моделях
- Microsoft Azure ML – це хмарна платформа, яка надає широкий спектр інструментів і послуг для створення, розгортання та керування моделями машинного навчання, включаючи можливості AutoML.

Це кілька прикладів того, як компанії використовують AutoML у різних галузях для автоматизації побудови моделей і налаштування гіперпараметрів, що дозволяє дослідникам обробки даних зосередитися на виборі та оцінці моделі.

AutoML автоматизує процес створення та налаштування моделей машинного навчання. Цей метод використовує алгоритми для пошуку найкращої моделі та гіперпараметрів, а не покладається на досвід людини. AutoML включає підвищену ефективність і здатність обробляти великі обсяги даних. Це може

бути корисним у разі нестачі досвідчених практиків машинного навчання. Однак для AutoML також є обмеження. Інтерпретація результатів автоматизованого процесу пошуку може бути дорогою з обчислювальної точки зору та важкою. Крім того, практичне використання AutoML обмежене якістю даних і доступністю обчислювальних ресурсів. На практиці AutoML в основному використовується в промислових умовах для підвищення продуктивності та продуктивності моделювання в таких сценаріях, як зображення, мова, текст та інші форми даних.

2.2 Алгоритм Gradient Boosting для задач прогнозу енергоспоживання

Моделі прогнозування є одними з найбільш часто використовуваних моделей машинного навчання. Підвищення градієнта — це метод, який виділяється своєю швидкістю й точністю прогнозування, особливо з великими та складними наборами даних. Він більш відомий як Gradient Boosting Machine або GBM. У цій роботі спочатку обговоримо математичний апарат, що лежить в основі алгоритму посилення градієнта.

Вивчаючи машинне навчання, ви, мабуть, стикалися з таким терміном, який називається бустинг. Це найбільш неправильно інтерпретований термін у галузі Data Science. Принцип, що лежить в основі алгоритмів бустингу, полягає в тому, що спочатку ми створюємо модель на основі навчального набору даних, а потім будуємо другу модель, щоб виправити помилки, присутні в першій моделі.

Припустімо, що є n точок даних і 2 класи виходу (0 і 1). Ви хочете створити модель для визначення класу тестових даних. Тепер ми випадково вибираємо спостереження з навчального набору даних і передаємо їх у модель 1 (M_1). Ми також припускаємо, що спочатку всі спостереження мають однакову вагу, що означає однакову ймовірність бути обраними.

Пам'ятайте, що в техніках об'єднання слабкі учні об'єднуються, щоб створити сильну модель, тому тут $M_1, M_2, M_3, \dots, M_n$ — усі слабкі учні.

Оскільки $M1$ погано навчається, вона, безумовно, неправильно класифікує деякі спостереження. Тепер, перед подачею спостережень до $M2$, ми оновлюємо ваги спостережень, які неправильно класифіковані.

Це саме явище трапляється в методах посилення градієнта, коли спостереження неправильно класифіковано, його вага оновлюється, а для тих, які правильно класифіковані, їх ваги зменшуються. Збільшується ймовірність вибору неправильно класифікованого спостереження. Отже, у наступній моделі вибираються лише ті спостереження, які були неправильно класифіковані в моделі 1.

Подібним чином це відбувається з $M2$, неправильно класифіковані ваги знову оновлюються, а потім передаються до $M3$. Ця процедура продовжується до тих пір, поки помилки не будуть мінімізовані, і набір даних не буде правильно передбачений. Тепер, коли надходить нова точка даних (тестові дані), вона проходить через усі моделі (слабкі учні), і клас, який отримує найбільше голосів, є результатом для наших тестових даних.

G – це комплексний метод машинного навчання, який послідовно поєднує передбачення кількох слабких учнів, як правило, дерев рішень (рис. 2.2.). Він спрямований на покращення загальної ефективності прогнозування шляхом оптимізації вагових коефіцієнтів моделі на основі помилок попередніх ітерацій, поступового зменшення помилок прогнозування та підвищення точності моделі. Це найчастіше використовується для лінійної регресії.

Помилки відіграють головну роль у будь-якому алгоритмі машинного навчання. В основному існує два типи помилок: помилка зсуву та помилка дисперсії. Алгоритм посилення градієнта допомагає мінімізувати помилку зміщення моделі. Основна ідея цього алгоритму полягає в тому, щоб послідовно будувати моделі, а ці наступні моделі намагаються зменшити помилки попередньої моделі. Але як нам це зробити? Як зменшити помилку? Це робиться шляхом побудови нової моделі на основі помилок або залишків попередньої моделі.

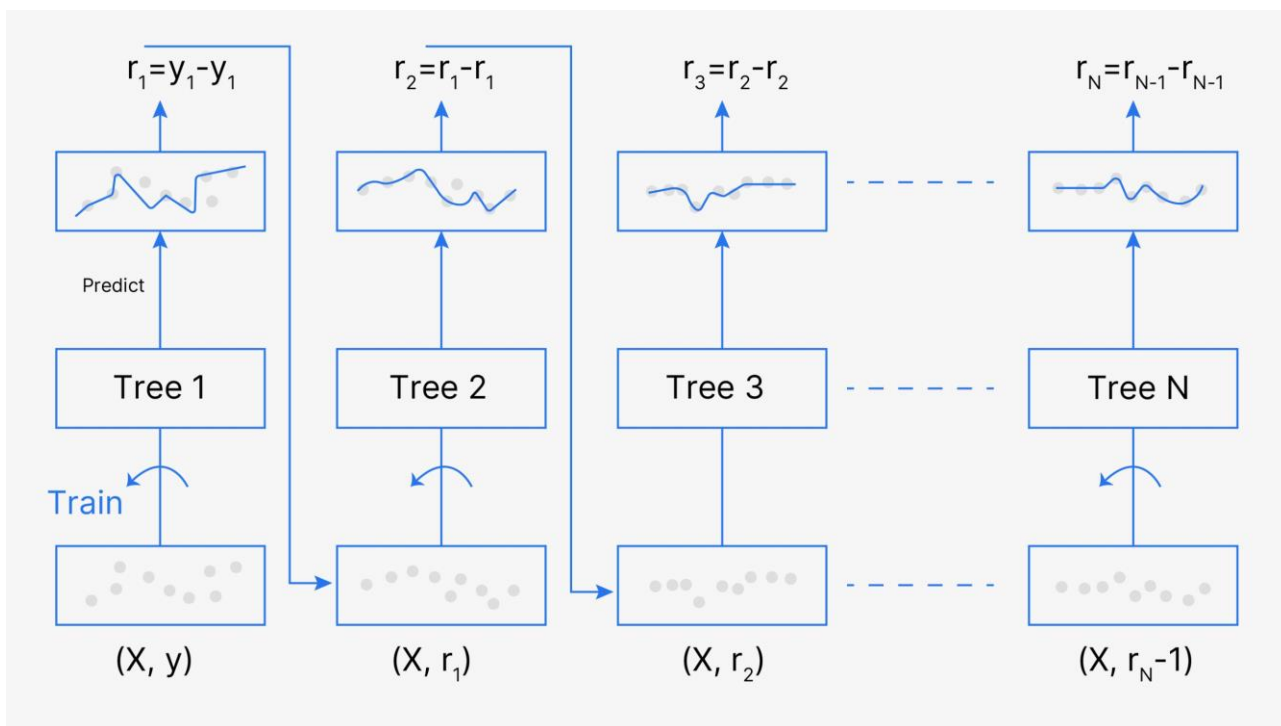


Рисунок 2.2 – Принцип методу Boosting

Коли цільовий стовпець безперервний, ми використовуємо градієнтний підсилювальний регресор, а коли це проблема класифікації, ми використовуємо градієнтний підсилювальний класифікатор. Єдина різниця між ними – «функція втрати». Метою тут є мінімізація цієї функції втрат шляхом додавання слабких учнів за допомогою градієнтного спуску. Оскільки він заснований на функції втрат, для задач регресії ми матимемо різні функції втрат, наприклад середньоквадратичну помилку (MSE), а для класифікації ми матимемо різні функції, як-от логарифм правдоподібності.

2.2.1 Приклад алгоритму Gradient Boosting (посилення градієнта)

Давайте зрозуміємо принцип алгоритму посилення градієнта за допомогою прикладу. Тут наш цільовий стовпець є безперервним, тому ми будемо використовувати регресор, що посилює градієнт.

Нижче наведено зразок із випадкового набору даних, де ми маємо передбачити ціну автомобіля на основі різних характеристик. Цільовий стовпець – ціна, а інші функції є незалежними (рис. 2.3).

Row No.	Cylinder Number	Car Height	Engine Location	Price
1	Four	48.8	Front	12000
2	Six	48.8	Back	16500
3	Five	52.4	Back	15500
4	Four	54.3	Front	14000

Рисунок 2.3 – Початковий набір даних

Крок 1: Створення базової моделі.

Першим кроком у посиленні градієнта є створення базової моделі для прогнозування спостережень у навчальному наборі даних. Для простоти ми беремо середнє значення цільового стовпця та припускаємо, що воно є прогнозованим значенням, як показано нижче (рис. 2.4).

Row No.	Cylinder Number	Car Height	Engine Location	Price	Prediction 1
1	Four	48.8	Front	12000	14500
2	Six	48.8	Back	16500	14500
3	Five	52.4	Back	15500	14500
4	Four	54.3	Front	14000	14500

Рисунок 2.4 – Перший крок прогнозування значень цільового атрибуту

Формально цей перший крок можна записати наступним чином:

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma). \quad (2.1)$$

Тут L — наша функція втрат, γ — наше прогнозоване значення, а $\arg \min$ означає, що ми повинні знайти прогнозоване значення/гаму, для яких функція втрат є мінімальною.

Оскільки цільовий стовпець неперервний, наша функція втрат буде:

$$L = \frac{1}{n} \sum_{i=0}^n (y_i - \gamma_i)^2 \quad (2.2)$$

Тут y_i — спостережене значення, а γ — прогнозоване значення.

Тепер нам потрібно знайти мінімальне значення γ , щоб ця функція втрат була мінімальною.

Тепер нам потрібно знайти мінімальне значення γ , щоб ця функція втрат була мінімальною. Для цього диференціюємо (2.2) по γ і прирівнюємо похідну до нуля.

$$\frac{dL}{d\gamma} = \frac{2}{2} \left(\sum_{i=0}^n (y_i - \gamma_i) \right) = - \sum_{i=0}^n (y_i - \gamma_i) \quad (2.3)$$

Пам'ятайте, що y_i — це наше спостережене значення, а γ_i — це наше прогнозоване значення. Додавши значення до наведеної вище формули, ми отримаємо:

$$L = \frac{1}{2}(12000 - \gamma)^2 + \frac{1}{2}(16500 - \gamma)^2 + \frac{1}{2}(15500 - \gamma)^2 + \frac{1}{2}(14000 - \gamma)^2$$

$$\frac{dL}{d\gamma} = \frac{2}{2}(12000 - \gamma)(-1) + \frac{2}{2}(16500 - \gamma)(-1) + \frac{2}{2}(15500 - \gamma)(-1) + \frac{2}{2}(14000 - \gamma)(-1)$$

Now $\frac{dL}{d\gamma} = 0$ and taking $(-)$ common

$$\Rightarrow -[12000 - \gamma + 16500 - \gamma + 15500 - \gamma + 14000 - \gamma] = 0$$

$$\Rightarrow [58000 - 4\gamma] = 0$$

$$\Rightarrow 58000 = 4\gamma$$

$$\Rightarrow \gamma = \frac{58000}{4} = 14500$$

У кінцевому підсумку ми перевищили середнє значення спостережуваної ціни автомобіля, тому я попросив вас взяти середнє значення цільового стовпця та припустити, що це ваш перший прогноз.

Отже, для $\gamma=14500$ функція втрат буде мінімальною, тому це значення стане нашим прогнозом для базової моделі.

Крок 2: Обчислення псевдозалишків.

Наступним кроком є обчислення псевдозалишків, які є різницею (спостережене значення – прогнозоване значення) (рис. 2.5).

Row No.	Cylinder Number	Car Height	Engine Location	Price	Prediction 1	Residual 1
1	Four	48.8	Front	12000	14500	-2500
2	Six	48.8	Back	16500	14500	2000
3	Five	52.4	Back	15500	14500	1000
4	Four	54.3	Front	14000	14500	-500

Рисунок 2.5 – Значення псевдозалишків

Цей крок можна записати так:

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad \text{for } i = 1, \dots, n. \quad (2.4)$$

Тут $F(x_i)$ — попередня модель, а m – кількість зроблених ДТ.

Ми просто беремо похідну функції втрат відносно прогнозованого значення, і ми вже обчислили цю похідну:

$$\frac{dL}{d\gamma} = -(y_i - \gamma_i) = -(Observed - Predicted)$$

Якщо ви бачите формулу залишків вище, ми бачимо, що похідна функції втрат помножена на від’ємний знак, тож тепер ми отримуємо:

$$(Observed - Predicted)$$

Передбачене значення тут є прогнозом, зробленим попередньою моделлю. У нашому прикладі прогноз, зроблений попередньою моделлю (початковий прогноз базової моделі), становить 14500, для обчислення залишків наша формула виглядає так: $(Observed - 14500)$.

Крок 3: Побудова моделі на основі обчислених залишків.

На наступному кроці ми побудуємо модель на основі цих псевдозалишків і зробимо прогнози. Оскільки ми хочемо мінімізувати ці залишки, мінімізація залишків зрештою покращить точність нашої моделі та потужність прогнозування. Отже, використовуючи Залишок як ціль і оригінальну функцію Кількість циліндрів, висоту циліндра та розташування двигуна, ми створимо нові прогнози. Зауважимо, що прогнози в цьому випадку будуть значеннями помилок, а не прогнозованими значеннями ціни автомобіля, оскільки наш цільовий стовпець тепер є похибкою.

Скажімо, $\mathbf{h}_m(\mathbf{x})$ – це наш DT, складений на цих залишках.

Крок 4. Обчислення вихідні дані дерева рішень.

На цьому кроці ми знаходимо вихідні значення для кожного листа нашого дерева рішень. Це означає, що може бути випадок, коли 1 листок отримує більше, ніж 1 залишок, тому нам потрібно знайти кінцевий вихід усіх листків. Щоб знайти результат, ми можемо просто взяти середнє значення всіх чисел у листі, не має значення, є лише 1 число чи більше 1.

Математично цей крок можна представити так:

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)). \quad (2.5)$$

Тут $\mathbf{h}_m(\mathbf{x}_i)$ — DT, зроблене на залишки, а \mathbf{m} — кількість DT. Коли $m=1$, ми говоримо про 1-й DT, а коли це « \mathbf{M} », ми говоримо про останній DT.

Вихідним значенням для листа є значення γ , яке мінімізує функцію втрат. Ліва сторона «Гамма» є вихідним значенням певного листа. Праворуч $[F_{m-1}(x_i) + \gamma h_m(x_i)]$ подібний до кроку 1, але тут різниця полягає в тому, що ми беремо попередні прогнози, тоді як раніше попередніх прогнозів не було.

2.2.2 Впровадження GBM за допомогою scikit-learn

Для реалізації GBM на наборі даних ми будемо використовувати набір даних оцінки доходу, який містить інформацію про особисте життя людини та

результат 50K або ≤ 50 . Набір даних можна знайти тут (<https://www.kaggle.com/lodetomasi1995/income-classification>)

Завдання тут полягає в тому, щоб класифікувати доходи особи, якщо отримати необхідні дані про її особисте життя.

Спочатку давайте імпортуємо всі необхідні бібліотеки (лістинг 2.1).

Лістинг 2.1 – Підключення бібліотек

```
# Import all relevant libraries
from sklearn.ensemble import GradientBoostingClassifier
import numpy as np
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn import preprocessing
import warnings
warnings.filterwarnings("ignore")
```

Тепер давайте прочитаємо набір даних і поглянемо на стовпці, щоб краще зрозуміти інформацію.

```
df = pd.read_csv('input_data.csv')
df.head()
```

Результат читання – на рисунку 2.6.

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country	income
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	$\leq 50K$
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	$\leq 50K$
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	$\leq 50K$
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	$\leq 50K$
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	$\leq 50K$

Рисунок 2.6 – Фай даних

Тепер для навчання та тестування нашої моделі дані потрібно розділити на навчальні та тестові набори. Ми також будемо масштабувати дані, щоб вони лежали між 0 і 1.

```
# Split dataset into test and train data
X_train, X_test, y_train, y_test =
train_test_split(df.drop('income', axis=1), df['income'],
test_size=0.2)
```

Тепер давайте продовжимо визначення класифікатора посилення градієнта разом із його гіперпараметрами. Далі ми впишемо цю модель у навчальні дані:

```
# Define Gradient Boosting Classifier with hyperparameters

gbc=GradientBoostingClassifier(n_estimators=500,
learning_rate=0.05,random_state=100, max_features=5)

# Fit train data to GBC

gbc.fit(X_train, y_train)
```

Модель навчено, і тепер ми також можемо спостерігати за результатами.

Нижче ви можете побачити матрицю помилок моделі, яка дає звіт про кількість класифікацій і неправильних класифікацій.

```
print(confusion_matrix(y_test, gbc.predict(X_test)))
[[4107  801]
 [ 533 4195]]
```

Кількість неправильних класифікацій за допомогою класифікатора Gradient Boosting Classifier становить 1334 у порівнянні з 8302 правильними класифікаціями. Модель показала себе гідно. Точність становить 86%, що досить добре, але її можна покращити, налаштувавши гіперпараметри або обробивши дані для видалення викидів. Однак це дає нам основну ідею посилення градієнта та принципи його роботи.

Точність становить 86%, що досить добре, але її можна покращити, налаштувавши гіперпараметри або обробивши дані для видалення викидів. Однак це дає нам основну ідею посилення градієнта та принципи його роботи.

3 ПРАКТИЧНА ЧАСТИНА РОБОТИ

3.1 Загальний опис роботи програми

Далі застосуємо описані вище засоби та методи для покращення процесу прогнозування щоденних рівнів енергоспоживання шляхом перетворення набору даних часових рядів у табличний формат за допомогою бібліотек з відкритим кодом. Ми досліджуємо застосування популярної багатокласової моделі класифікації та використовуємо AutoML із Cleanlab Studio, щоб значно підвищити нашу точність поза вибіркою.

Ключовий висновок полягає в тому, що ми можемо використовувати більш загальні методи для моделювання набору даних часового ряду, перетворюючи його на табличну структуру, і навіть знаходити покращення в спробі передбачити дані цього часового ряду.

На високому рівні ми:

- встановлюємо базову точність, пристосувавши модель прогнозування Prophet до наших даних часових рядів;
- перетворюємо наші дані часового ряду в табличний формат, використовуючи бібліотеки характеристик із відкритим вихідним кодом, а потім покажемо, що може перевершити нашу модель Prophet за допомогою стандартного підходу багатокласової класифікації (підсилення градієнта) за рахунок зменшення помилки прогнозування на 67% (збільшення на 38% вихідні відсоткові пункти в точності поза вибіркою);
- використання рішення AutoML для багатокласової класифікації призвело до зменшення помилки передбачення на 42% (збільшення на 8% у необроблених відсоткових пунктах точності поза вибіркою) порівняно з нашою моделлю посилення градієнта та призвело до зменшення помилки передбачення на 81% (збільшення на 46% у необроблених відсоткових пунктах у точності поза вибіркою) порівняно з нашою моделлю прогнозування Prophet.

3.1.1 Аналіз вхідного набору даних

Скористаємося набором даних про погодинне споживання електроенергії компанією, доступним за адресою [12]. Звичайно, що даний програмний продукт може бути адаптований до аналогічних даних будь-якого українського населеного пункту.

Дані представляють погодинне споживання енергії PJM (у мегаватах) на погодинній основі. PJM Interconnection LLC (PJM) є регіональною організацією з передачі (RTO) у Сполучених Штатах. Це частина мережі Східного з'єднання, яка керує системою електропередачі, яка обслуговує багато штатів.

Давайте подивимося на наш набір даних. Дані містять один стовпець дати й часу (тип об'єкт) і стовпець Споживання енергії в мегаватах (тип float64), який ми намагаємося спрогнозувати як дискретну змінну (що відповідає квартилю рівнів погодинного енергоспоживання). Наша мета полягає в тому, щоб навчити модель прогнозування часових рядів, щоб мати можливість прогнозувати завтрашній щоденний рівень енергоспоживання, потрапляючи на 1 із 4 рівнів: низький, нижче середнього, вище середнього або високого (ці рівні були визначені на основі квартилів загального добового розподілу споживання). Спочатку ми продемонструємо, як застосувати до цієї проблеми такі методи прогнозування часових рядів, як Prophet, але вони обмежені певними типами моделей ML, які підходять для даних часових рядів. Далі ми демонструємо, як переформувати цю задачу в стандартну багатокласову проблему класифікації, до якої ми можемо застосувати будь-яку модель машинного навчання, і показуємо, як ми можемо отримати кращі прогнози за допомогою потужного контрольованого ML.

Спочатку ми перетворюємо ці дані в середнє споживання енергії на щоденному рівні та перейменовуємо стовпці у формат, який очікує модель прогнозування Prophet. Ці реальні добові рівні споживання енергії перетворюються в квартилі, що є значенням, яке ми намагаємося передбачити. Нижче наведено наші тренувальні дані разом із квартилем, до якого потрапляє

кожен рівень денного споживання енергії. Квартилі обчислюються з використанням навчальних даних, щоб запобігти витoku даних.

	ds	y	quartile
0	2002-01-01	31080.739130	2
1	2002-01-02	34261.541667	3
2	2002-01-03	34511.875000	3
3	2002-01-04	33715.458333	3
4	2002-01-05	30405.125000	2
...
4842	2015-04-05	24577.500000	1
4843	2015-04-06	26996.666667	1
4844	2015-04-07	27177.833333	1
4845	2015-04-08	29136.041667	2
4846	2015-04-09	30535.291667	2

4847 rows x 3 columns

Рисунок 3.1 – Вхідний тренувальний набір даних з диференціацією цільового атрибуту по кuartилях

З цього ж набору даних виберемо тестовий набір, на котрому перевіримо точність моделі машинного навчання.

3.2 Тренування та оцінка моделі прогнозування Prophet

Як видно на зображеннях вище, ми використаємо граничну дату 2015-04-09, щоб завершити діапазон наших навчальних даних і почати наші тестові дані 2015-04-10. Ми обчислюємо кuartильні пороги нашого щоденного споживання енергії, використовуючи ТІЛЬКИ дані тренувань. Це дозволяє уникнути витoku даних - використання даних поза вибіркою, які будуть доступні лише в майбутньому.

Далі ми спрогнозуємо щоденний рівень енергоспоживання PJME (у МВт) протягом періоду наших тестових даних і представимо прогнозовані значення як

дискретну змінну. Ця змінна показує, до якого квартиля потрапляє рівень денного споживання енергії, представлений категорично як 1 (низький), 2 (нижче середнього), 3 (вище середнього) або 4 (високий). Для оцінки ми будемо використовувати функцію `accuracy_score` від `scikit-learn`, щоб оцінити продуктивність наших моделей. Оскільки ми формулюємо проблему таким чином, ми можемо оцінити прогнози нашої моделі на наступний день (і порівняти майбутні моделі), використовуючи точність класифікації (лістинг 3.1).

```
import numpy as np
from prophet import Prophet
from sklearn.metrics import accuracy_score

# Initialize model and train it on training data
model = Prophet()
model.fit(train_df)

# Create a dataframe for future predictions covering the test period
future = model.make_future_dataframe(periods=len(test_df), freq='D')
forecast = model.predict(future)

# Categorize forecasted daily values into quartiles based on the thresholds
forecast['quartile'] = pd.cut(forecast['yhat'], bins = [-np.inf] + list(quartiles) + [np.inf], labels=[1, 2, 3, 4])

# Extract the forecasted quartiles for the test period
forecasted_quartiles = forecast.iloc[-len(test_df):]['quartile'].astype(int)

# Categorize actual daily values in the test set into quartiles
test_df['quartile'] = pd.cut(test_df['y'], bins=[-np.inf] + list(quartiles) + [np.inf], labels=[1, 2, 3, 4])
actual_test_quartiles = test_df['quartile'].astype(int)

# Calculate the evaluation metrics
prophet_accuracy = accuracy_score(actual_test_quartiles, forecasted_quartiles)

# Print the evaluation metrics
print(f'Accuracy: {prophet_accuracy:.4f}')
```

Отримаємо точність поза вибіркою досить низька – 43%. Моделюючи часові ряди таким чином, ми обмежуємося використанням лише моделей прогнозування часових рядів (обмеженої підмножини можливих моделей ML). У наступному розділі ми розглянемо, як ми можемо більш гнучко моделювати ці дані шляхом перетворення часових рядів у стандартний табличний набір даних за допомогою відповідної характеристики. Після того, як часовий ряд буде перетворено на стандартний табличний набір даних, ми зможемо використовувати будь-яку контрольовану модель ML для прогнозування цих щоденних даних споживання енергії.

3.3 Перетворення даних часових рядів у табличні дані за допомогою функції

Тепер ми перетворюємо дані часових рядів у табличний формат і представляємо дані за допомогою бібліотек з відкритим кодом `sktime`, `tsfresh` і `tsfel`. Використовуючи подібні бібліотеки, ми можемо отримувати широкий спектр функцій, які фіксують базові шаблони та характеристики даних часових рядів. Це включає статистичні, часові та, можливо, спектральні характеристики, які забезпечують повний знімок поведінки даних у часі. Розбиваючи часові ряди на окремі ознаки, стає легше зрозуміти, як різні аспекти даних впливають на цільову змінну.

`TSFreshFeatureExtractor` — це інструмент вилучення функцій із бібліотеки `sktime`, який використовує можливості `tsfresh` для вилучення відповідних функцій із даних часових рядів. `tsfresh` розроблено для автоматичного обчислення великої кількості характеристик часових рядів, що може бути дуже корисним для розуміння складної часової динаміки. Для нашого випадку використання ми використовуємо мінімальний і суттєвий набір функцій із нашого `TSFreshFeatureExtractor` для представлення наших даних.

`tsfel`, або бібліотека вилучення ознак часового ряду, пропонує повний набір інструментів для вилучення ознак із даних часового ряду. Ми використовуємо попередньо визначену конфігурацію, яка дозволяє створювати багатий набір функцій (наприклад, статистичні, часові, спектральні) з даних часових рядів споживання енергії, фіксуючи широкий діапазон характеристик, які можуть бути доречними для нашого завдання класифікації (лістинг 3.2).

Лістинг 3.2 – Перетворення часових рядів у табличну форму

```

import tsfel
from sktime.transformations.panel.tsfresh import TSFreshFeatureExtractor

# Define tsfresh feature extractor
tsfresh_trafo = TSFreshFeatureExtractor(default_fc_parameters="minimal")

# Transform the training data using the feature extractor
X_train_transformed = tsfresh_trafo.fit_transform(X_train)

# Transform the test data using the same feature extractor
X_test_transformed = tsfresh_trafo.transform(X_test)

# Retrieves a pre-defined feature configuration file to extract all available features
cfg = tsfel.get_features_by_domain()

# Function to compute tsfel features per day
def compute_features(group):
    # TSFEL expects a DataFrame with the data in columns, so we transpose the input group
    features = tsfel.time_series_features_extractor(cfg, group, fs=1, verbose=0)
    return features

# Group by the 'day' level of the index and apply the feature computation
train_features_per_day = X_train.groupby(level='Date').apply(compute_features).reset_index(drop=True)
test_features_per_day = X_test.groupby(level='Date').apply(compute_features).reset_index(drop=True)

# Combine each featurization into a set of combined features for our train/test data
train_combined_df = pd.concat([X_train_transformed, train_features_per_day], axis=1)
test_combined_df = pd.concat([X_test_transformed, test_features_per_day], axis=1)

# Filter out features that are highly correlated with our target variable
column_of_interest = "PJME_MW_mean"
train_corr_matrix = train_combined_df.corr()
train_corr_with_interest = train_corr_matrix[column_of_interest]
null_corrs = pd.Series(train_corr_with_interest.isnull())
false_features = null_corrs[null_corrs].index.tolist()

columns_to_exclude = list(set(train_corr_with_interest[abs(train_corr_with_interest) > 0.8].index.tolist() + false_features)
                           - {column_of_interest})

# Filtered DataFrame excluding columns with high correlation to the column of interest
X_train_transformed = train_combined_df.drop(columns=columns_to_exclude)
X_test_transformed = test_combined_df.drop(columns=columns_to_exclude)

```

Результат показано на наступному рисунку 3.2.

Feature Extraction: 100%	4817/4817	[00:00<00:00, 7909.95it/s]
Feature Extraction: 100%	1205/1205	[00:00<00:00, 9343.20it/s]

Рисунк 3.2 – Перетворення часових рядів у табличну форму

Далі ми очищаємо наш набір даних, видаляючи функції, які показали високу кореляцію (вище 0,8) з нашою цільовою змінною — середньодобовим рівнем споживання енергії — і ті, що мають нульову кореляцію. Характеристики високої кореляції можуть призвести до переобладнання, коли модель добре працює на навчальних даних, але погано на невидимих даних. Нуль-корельовані ознаки, з іншого боку, не надають жодної цінності, оскільки їм бракує визначеного зв'язку з ціллю.

Виключаючи ці функції, ми прагнемо покращити можливість узагальнення моделі та гарантувати, що наші прогнози базуються на збалансованому та значущому наборі вхідних даних (лістинг 3.3).

Лістинг 3.3 – Означення функції фільтрації значущих даних

```
# Filter out features that are highly correlated with our target variable
column_of_interest = "PJME_MW__mean"
train_corr_matrix = train_combined_df.corr()
train_corr_with_interest = train_corr_matrix[column_of_interest]
null_corrs = pd.Series(train_corr_with_interest.isnull())
false_features = null_corrs[null_corrs].index.tolist()

columns_to_exclude = list(set(train_corr_with_interest[abs(train_corr_with_interest) > 0.8].index.tolist() + false_features
                             columns_to_exclude.remove(column_of_interest)

# Filtered DataFrame excluding columns with high correlation to the column of interest
X_train_transformed = train_combined_df.drop(columns=columns_to_exclude)
X_test_transformed = test_combined_df.drop(columns=columns_to_exclude)
```

Тепер у нас є 73 функції, які були додані з використаних нами бібліотек характеристик часових рядів. Мітка, яку ми збираємося передбачити на основі цих характеристик, — це рівень споживання енергії наступного дня. Перші 5 рядків навчальних даних, які нещодавно представлені в табличному форматі (рисунок

	PJME_MW__standard_deviation	PJME_MW__variance	PJME_MW_Centroid	PJME_MW_Entropy	PJME_MW_FFT mean coefficient_0	PJME_MW_FFT mean coefficient_1
0	4097.961271	1.679329e+07	12.727435	1.0	5.207144e+06	1.736704e+08
1	3718.008117	1.382358e+07	12.554067	1.0	3.425893e+06	1.564219e+08
2	3241.304817	1.050606e+07	12.395692	1.0	2.600067e+06	1.015409e+08
3	2259.371710	5.104761e+06	12.204000	1.0	4.219601e+04	6.502275e+07
4	3250.463504	1.056551e+07	12.751234	1.0	7.678627e+05	2.307445e+08

5 rows x 73 columns

Рисунок 3.3 – Тренувальний набір даних

Важливо зазначити, що ми використали найкращу практику застосування процесу створення функцій окремо для навчальних і тестових даних, щоб уникнути витоку даних (і затримані тестові дані є нашими останніми спостереженнями).

Крім того, ми обчислюємо наше дискретне значення квартиля (використовуючи квартилі, які ми спочатку визначили), використовуючи

наведений нижче код, щоб отримати наші енергетичні мітки тренування/тесту, якими є наші `y_labels` (лістинг 3.4).

Лістинг 3.4 – Функція перетворення цільового стовпця у кuartилі

```
# Define a function to classify each value into a quartile
def classify_into_quartile(value):
    if value < quartiles[0]:
        return 1
    elif value < quartiles[1]:
        return 2
    elif value < quartiles[2]:
        return 3
    else:
        return 4

y_train = X_train_transformed["PJME_MW_mean"].rename("daily_energy_level")
X_train_transformed.drop("PJME_MW_mean", inplace=True, axis=1)

y_test = X_test_transformed["PJME_MW_mean"].rename("daily_energy_level")
X_test_transformed.drop("PJME_MW_mean", inplace=True, axis=1)

energy_levels_train = y_train.apply(classify_into_quartile)
energy_levels_test = y_test.apply(classify_into_quartile)
```

3.4 Навчання та оцінка моделі GradientBoostingClassifier на основі представлених табличних даних

Використовуючи наш спеціальний набір табличних даних, ми можемо застосувати будь-яку модель ML для прогнозування майбутніх рівнів енергоспоживання. Тут ми скористаємося моделлю Gradient Boosting Classifier (GBC), яку вибирають більшість дослідників даних, які працюють із табличними даними.

Наша модель GBC створена з модуля `sklearn.ensemble` і налаштована за допомогою певних гіперпараметрів, щоб оптимізувати її продуктивність і уникнути перенавчання (лістинг 3.5).

Лістинг 3.5 – Навчання моделі передбачення енергоспоживання

```
from sklearn.ensemble import GradientBoostingClassifier

gbc = GradientBoostingClassifier(
    n_estimators=150,
    learning_rate=0.1,
    max_depth=4,
    min_samples_leaf=20,
    max_features='sqrt',
    subsample=0.8,
    random_state=42
)
```

```
gbc.fit(X_train_transformed, energy_levels_train)

y_pred_gbc = gbc.predict(X_test_transformed)
gbc_accuracy = accuracy_score(energy_levels_test, y_pred_gbc)
print(f'Accuracy: {gbc_accuracy:.4f}')
```

В результаті отримана точність 81% значно краща, ніж наші попередні результати моделі Prophet.

3.5 Використання AutoML для оптимізації

Тепер, коли ми побачили, як відобразити задачу часових рядів і переваги застосування потужних моделей машинного навчання, таких як Gradient Boosting, виникає природне запитання: яку модель машинного навчання слід застосувати? Звичайно, ми могли б експериментувати з багатьма моделями, налаштовувати їхні гіперпараметри та складати їх разом. Простіше рішення — дозволити AutoML виконувати все це за нас.

Тут ми використаємо просте рішення AutoML, надане в Cleanlab Studio , яке передбачає нульову конфігурацію. Ми просто надаємо наш табличний набір даних, і платформа автоматично навчає багато типів керованих моделей ML (зокрема, посилення градієнта), налаштовує їхні гіперпараметри та визначає, які моделі найкраще об'єднати в єдиний предиктор. Ось увесь код, необхідний для навчання та розгортання контрольованого класифікатора AutoML (лістинг 3.6).

Лістинг 3.6 – Використання AutoML

```
from cleanlab_studio import Studio

studio = Studio()
studio.create_project(
    dataset_id=energy_forecasting_dataset,
    project_name="ENERGY-LEVEL-FORECASTING",
    modality="tabular",
    task_type="multi-class",
    model_type="regular",
    label_column="daily_energy_level",
)

model = studio.get_model(energy_forecasting_model)
y_pred_automl = model.predict(test_data, return_pred_proba=True)

y_pred_automl_cleanlab = pd.read_csv("quartile-multiclass-pjme-testing-data_pred_probs.csv")
y_pred_automl_cleanlab = y_pred_automl_cleanlab["Suggested Label"]
```

На рисунку 3.4 нижче ми можемо побачити оцінки моделі на платформі AutoML, де показано всі різні типи моделей ML, які були автоматично підігнані та оцінені (включно з моделями посилення градієнта), а також ансамблевий предиктор, створений шляхом оптимального поєднання їхніх прогнозів.

Model Type	Accuracy
Ensemble Predictor	0.954121
Neural Network (fast.ai)	0.947062
Neural Network (MLP)	0.945609
Gradient Boosting (LightGBM)	0.920490
Gradient Boosting 0	0.915508
Gradient Boosting 1	0.912186
Gradient Boosting (LightGBM, num_leaves=128)	0.905128
Gradient Boosting (ExtraTrees Splits)	0.904712
Nearest Neighbors (Distance-Weighted)	0.875234
Random Forest (Entropy Criterion)	0.873365

Рисунок 3.4 – AutoML для різних типів використаних моделей

Зробивши висновок на основі наших тестових даних, щоб отримати прогнози рівня споживання енергії наступного дня, ми бачимо, що точність тесту становить 89%, що на 8 % краще порівняно з нашим попереднім підходом Gradient Boosting (рис. 3.5).

```
automl_accuracy = accuracy_score(energy_levels_test, y_pred_automl_cleanlab)
print(f'Accuracy: {automl_accuracy:.4f}')
```

Accuracy: 0.8880

Рисунок 3.5 – Точність моделі AutoML

Отже, AutoML підвищує точність прогнозу щоденних даних про рівень енергоспоживання.

4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Охорона праці та її актуальність в ІТ-сфері

Для підвищення ефективності системи управління охорони праці (СУОП) дуже важлива роль належить формуванню і розвитку інформаційної культури фахівців ІТ-технологій, яка впливає на удосконалення інформаційного контуру сучасних підприємств, дозволяє створювати надійні прогнози щодо стану умов праці, показників здоров'я та працездатності, виробничого травматизму і професійної захворюваності, визначати політику розвитку підприємств, установ та організацій на основі різноманітних стратегій охорони праці (інноваційні, маркетингові, інвестиційні, фінансові, технологічні, диверсифікаційні). Поряд з інформаційною культурою важливо використовувати в рамках СУОП «трикутник» її складових: правову, організаційну, управлінську.

В управлінні охороною праці потрібно реалізувати основні положення, окремі теоретико-методологічні підходи інформаційного менеджменту. Головну роль та відповідальність за стан СУОП мають нести фахівці служби охорони праці сучасного підприємства.

Сучасне суспільство називають постіндустріальним, постеконічним, інформаційним, оскільки йдеться про багатосторонні і кардинальні зміни у розвитку цивілізації.

На постіндустріальному етапі розвитку суспільства вирішальним фактором стає інформація. Її домінування ініціювала науково-технічна революція, яку ще іменують інформаційною, оскільки нею охоплена будь-яка інтелектуальна діяльність, починаючи з інформаційних образів штучного інтелекту у нових технологіях, економіки, і продовжуючи інформатизацією суспільства в умовах світової глобалізації науки й освіти тощо.

Інформаційні технології розглядаються як потужний важіль економічного зростання України. Для цього необхідні значні стратегічні інвестиції у

комп'ютерну та комунікаційну інфраструктуру, програми досліджень і розробок, освітню галузь [26].

Під інформаційною культурою розуміють сукупність, складову НІТ (новітні інформаційні технології), технологічну, правову, психологічну, соціологічну та ергономічну підсистеми, що сприяють спрямованому впливу на протікання соціальних процесів у суспільстві, колективі і вихованню свідомого відношення людини до праці, виконання прав та обов'язків [27].

Поняття інформаційної культури виникло в процесі активізації дослідницької уваги до механізмів інформаційного обміну у зв'язку зі значним підвищення ролі інформації в соціокультурних процесах суспільства, яке розглядають як інформаційне суспільство знань, де в центрі знаходяться інформаційні технології.

Робота з інформацією та інформаційна культура в цілому є одним з найважливіших компонентів спроб компанії управляти змінами. Є три принципові причини, в силу яких сьогодні необхідно дбати про інформаційну культуру компанії.

По-перше, вона все більше і більше стає найважливішою частиною загальної організаційної (корпоративної) культури компанії. Все більше компаній розуміють необхідність перетворень, орієнтованих на задоволення очікувань споживача. Щоб сьогодні впливати на майбутнє, потрібно уявляти собі на що вона буде схожа. А для цього потрібно працювати з різноманітною діловою, професійною, технологічною, соціальною, ринковою та політичною інформацією.

По-друге, інформаційні технології роблять можливим створення в компаніях комп'ютерних мереж, за допомогою яких йде спілкування між менеджерами, але важливо знати, як люди використовують цю інформацію. Саме по собі створення такої мережі з усіма її робочими станціями і мультимедійними можливостями не гарантує того, що інформація буде використовуватися більш розумно і більш ефективно.

По-третє, для різних функціональних служб, підрозділів та робочих груп сучасних підприємств в сфері охорони праці інформаційна культура різна, а це означає відмінність методологічних підходів до процесів усвідомлення, збору, організації, обробки, поширення і використання інформації. Тому багато менеджерів погодяться з тим, що корпоративна інформаційна культура важлива для вироблення різних стратегій охорони праці та запровадження відповідних заходів з її вдосконалення.

Для деяких галузей, таких як розробка програмного забезпечення, інформаційна культура є необхідною умовою виживання, тому що зміна технологій в розробці програмного забезпечення відбувається кожні 6-8 місяців, а інвестиції на підготовку персоналу і освоєння нової технології величезні і у великих компаніях варіюються від 1,5 до 2 млрд. доларів на рік [27].

Аналіз свідчить, що інформатизація та інтеграція комунікаційного простору України сприяє різкому підвищенню інформаційної та професійної компетентності, ділової активності, стимулюванню конкуренції, створенню інноваційних підприємств та організацій, нових робочих місць, зниженню витрат на утримання управлінського апарату [25].

Поряд із задачами і здобутками окреслилися негативи використання інформаційних технологій:

1) надмірне інформаційне навантаження, суть якого полягає у тому, що кількість корисної інформації, яка надходить до мережі, перевищує психофізіологічні можливості її сприйняття людиною;

2) велика кількість інформації, яка сприймається, але не є корисною для фахівців в даний момент;

3) інформаційний голод, причиною якого є саме надлишок інформації, викликаний інформаційним перенавантаженням;

4) «інформоманія» як хвороба людини, яка робить останню знеособленою, залежною від перебування в інформаційному просторі і роботи з комп'ютером і чому вона віддає перевагу, уникаючи «живого» спілкування з людьми;

5) поява «кіберспільнот», що за своїми соціокультурними характеристиками набагато ближчі до представників інших культур у глобальному інформаційному просторі, ніж до своєї етнонаціональної спільноти чи решти населення, не охопленого Інтернетом;

б) індивідуалізм і дегуманізація способу життя «мешканців» Інтернету – відсутність готовності ділитися своїми знаннями.

Слід розуміти, що комп'ютерні технології, а особливо їх мережі істотно впливають на життєдіяльність людини, припускаючи глобалізацію і технократизацію суспільства. Але в ще більшій мірі цей вплив поширюється безпосередньо на центральну нервову систему, яка звикає працювати в дуже інтенсивному режимі багатозадачності, де вже переважають не тривалі логічні роздуми, а інтуїтивно-реактивні ланцюжки розумових формулювань у зв'язку з величезним обсягом оброблюваної щодня інформації, кількість якої зростає за експоненціальною швидкістю. Виникає припущення, що саме збільшення обсягу інформації та прискорення її обробки людиною може згубно вплинути на розвиток розумових здібностей людини.

Аналіз продуктивності розумової праці в найбільших за чисельністю фахівців ІТ-фірм показав, що велике значення з точки зору впливу на її результати має організаційна (корпоративна) культура. В цьому напрямі влаштовуються різні тимблдинги, заходи, тренінги для розвитку персоналу. Також кожен керівник повинен добре розуміти свого співробітника, що саме для нього важливо, що його мотивує. Важливо відвести потрібну роль відповідному співробітнику, щоб він виконував ті завдання, які йому цікаві.

На подібних тренінгах в тому числі повинна розглядатися інформаційна культура працівника, в освоєнні, володінні, мотивуванні, застосуванні, перетворенні інформації із застосуванням сучасних інформаційних технологій і використанням цих умінь в навчанні з охорони праці і в подальшій професійній діяльності. Особливо вони будуть корисні, як доповнення до існуючих інструктажів з охорони праці на підприємстві, або як контроль психологічного стану та взаємовідносин у колективі.

Інформаційна культура як інтегративне утворення абсолютно не зводиться до розрізнених знань, вмінь та навичок роботи за комп'ютером. Вона передбачає інформативну спрямованість цілісної особистості, яка володіє мотивацією до застосування і засвоєння нових даних. Інформаційну культуру можна розглядати, як одну з граней особистісного розвитку промислових робітників. Це шлях універсалізації якостей людини.

Оволодіння інформаційною культурою сприяє реальному розумінню особистістю свого місця, себе і своєї ролі у виробничому колективі. Вона має сприяти формуванню нового покоління фахівців інформаційного суспільства, який повинен володіти наступними навичками: виділення релевантної, значущої інформації, диференціації вихідних даних, розробки інформативних критеріїв її оцінки інформації, вміння використовувати її в рамках СУОП.

Сьогодні продовжує діяти стратегічне правило «Можливості комп'ютерної техніки обмежені тільки нашими уявленнями» [25].

4.2 Шкідлива дія шуму та вібрації і захист від неї

Для запобігання шкідливої дії шуму і вібрації на організм працюючих проводяться технічні, організаційні і медикопрофілактичні заходи.

Одним з основних технічних заходів є зменшення при експлуатації та на стадії проектування, конструювання обладнання причин шуму і вібрації в самому джерелі утворення. Досягають цього завдяки використанню раціональної конструкції обладнання, заміни ударної дії деталей і машин коливальною, з'єднання елементів гнучкими зв'язками, врівноважування обертових частин механізмів, заміни металевих деталей пластмасовими, забезпечення різних власних частот коливань механізму з частотою збуджуючої сили. Аеродинамічний шум може бути зменшений застосуванням глушників та повітропроводів зі змінним перерізом. Шум трансформаторів (електромагнітний шум) знижується, якщо застосувати листи заліза як складових осердя трансформатора з малою магнітострикцією, серцевини.

Якщо неможливо ізолювати чи знизити шум і вібрацію самого джерела, потрібно:

- ізолювати джерело шуму або вібрації від навколишнього середовища засобами вібро- та звукоізоляції;
- раціонально планувати виробничі приміщення, що мають інтенсивні джерела шуму;
- збільшувати звукопоглинання внутрішніх поверхонь приміщення шляхом звукопоглинальних покриттів.

Принцип роботи звукоізоляційних екранів оснований на відбиванні звукової хвилі від різних екранів, стін, кожухів обладнання. Шумливі агрегати слід закривати звукоізоляційними кожухами з виводом назовні органів керування та контрольних приладів. Звукоізоляційні екрани виготовляють з металу, деревини, пластмаси та інших щільних матеріалів. Екрани зсередини покривають звукопоглинаючими матеріалами (скловатою пінополіуретаном), а по периметру кожуха – віброізоляційними підкладками (гума).

Вихідними даними для розрахунку параметрів необхідного екрану є спектр шуму, який необхідно ослабити, кількість екранів, через які проходить шум, їх площа, акустичні характеристики приміщення. За розрахованими значеннями необхідної звукової ізоляційної здатності екрану підбирається матеріал конструкції й екрану.

Принцип звукопоглинання оснований на явищі трансформації коливальної енергії звуку в теплову через втрати при терті. Найбільші втрати при терті мають пористі, волокнисті і перфоровані матеріали: поролон, пемзолітові і деревоволокнисті плити тощо. Енергія звукової хвилі переходить у теплову енергію, причому, ефект звукоізоляції збільшується з ростом частоти звукової хвилі. Звукопоглинаючими матеріалами оббивають стелі, стіни. Щоб одержати ефективну звукоізоляцію, найбільш доцільно застосовувати багат шарові огороження з м'якими прошарками (мінеральна вата).

Важливим технічним рішенням у забезпеченні виробничих умов є вдосконалення ручних віброінструментів. Для цього використовують

віброгасіння, змінюють ударний вузол, проводять балансування частин, що обертаються.

Послаблення локальної вібрації і передачі вібрації на підлогу і сидіння досягається засобами віброізоляції і вібропоглинання, застосуванням пружинних і гумових амортизаторів, прокладок тощо. Для обмеження поширення вібрацій через ґрунт, між фундаментом і ґрунтом залишають повітряні проміжки, які називаються акустичними розривами.

В останні роки знаходять застосування динамічні віброгасники, в яких створюються вібрації, що співпадають по частоті і протилежні по фазі вібрації машини, коливання якої необхідно зменшити.

До організаційних заходів по боротьбі з шумом та вібрацією на виробництві відносяться: впровадження раціонального режиму праці і відпочинку, обмеження часу роботи при використанні ручного інструменту, який створює вібрацію.

Глушники звуку застосовуються для зменшення шуму аеродинамічних установок (вентиляторів, пневмоінструментів, газотурбінних, дизельних, компресорних установок). Вони поділяються на активні, які поглинають звукову енергію, що на них поступила, і реактивні, які відбивають цю енергію. Потужні джерела шуму як правило розміщують в окремих приміщеннях, які віддалені від постійних робочих місць. Ізоляційні кабінки або екрани застосовують як екрани робочих місць для зменшення зовнішніх шумів.

Якщо не вдається зменшити рівень шуму і вібрації на робочому місці до нормативних значень та необхідно використовувати засоби індивідуального захисту: рукавиці, взуття, навушники, м'які шоломи, які зменшують рівень звукового тиску на 40-50 дБ.

У процесі виробництв, експлуатації і зберігання радіоелектронних засобів можуть виникати механічні і динамічні дії, що характеризуються широким діапазоном частот коливань, а також амплітудою, прискоренням і часом дії. Рівень механічних дій визначається умовами транспортування й експлуатації.

Необхідно розрізняти два види механічних дій: удари і вібрації. Удар виникає, коли апаратура отримує швидку зміну прискорення (піддаються удару входи кабелів, джгути, резистори, конденсатори, напівпровідникові діоди і тріоди, силові трансформатори, дроселі тощо). Вібрації – довготривалі знаковмінні процеси, які впливають на роботу апаратури при безпосередньому контакті з джерелом коливань або через повітряне середовище.

У результаті дії вібрацій і удару можуть бути наступні ушкодження апаратури: порушення герметичності через псування паяльних, зварних і клеєних швів і появи тріщин у метало-скляних спаях; повне руйнування корпусів або окремих їх частин через механічний резонанс або циклічну втому; обривання монтажних зв'язків, відшарування багат шарових друкованих плат, руйнування підставок; вихід з ладу електричних контактів; модуляція розмірів хвилеводних трактів; коаксіальних кабелів, конденсаторів змінної ємності, коливальних контурів, електровакуумних приладів, зміщення положення органів настроювання і управління.

Під впливом вібрацій може статись зміна параметрів напівпровідникових приладів, вольт амперних характеристик діодів, транзисторів. Все це призводить до руйнування конструкцій за рахунок явищ втоми. Радіоелектронна апаратура (РЕА) повинна мати віброміцність, вібростійкість, ударостійкість.

Захист РЕА здійснюється наступними групами методів:

- зменшується інтенсивність джерел вібрації шляхом балансування, зменшення зазорів, віброізоляції джерела вібрацій;
- зменшується величина дій, що передається апаратом шляхом віброізоляції, демпфірування, виключення резонансів, активного віброзахисту за допомогою ексцентриків, маятників, гіроскопів;
- використання найбільш добротні і жорсткі компоненти і вузли;
- застосовуються амортизатори.

Захист часом, захист віддалю, усунення джерела тепловиділення, теплоізоляція, охолодження гарячої поверхні, забезпечення тепловіддачі тіла людини та індивідуальні засоби захисту. Захист часом передбачає обмеження

часу перебування робітника в зоні дії інфрачервоного випромінювання. Потужність випромінювання можна знизити за рахунок конструкторських і технологічних рішень (зміною нагрівання виробів у нагрівальних пічках індукційним нагріванням та ін.) і за рахунок покриття поверхні, яка нагрівається, тепло ізолювальним матеріалом.

Якщо теплоізоляція неможлива, тоді захист від прямої дії інфрачервоного випромінювання здійснюється екрануванням. Екрани можуть бути прозорими, напівпрозорими і непрозорими. У свою чергу вони поділяються на тепловідбивальні, тепловідвідні та теплопоглинальні; стаціонарні і нестаціонарні.

Застосовують також прозору водяну завісу у вигляді суцільної тонкої водяної плівки. Вода є активним поглиначем інфрачервоного випромінювання.

Перегрівання людини попереджують раціональним режимом життя, режимом праці та гідро процедурами. Спецодяг виготовляється з незаймистого, стійкого до інфрачервоного випромінювання, м'якого і повітронепроникного матеріалу (тканина з металевим покриттям відбиває 90% інфрачервоного випромінювання). Для захисту очей застосовують світлофільтри зі спеціального жовто-зеленого або синього скла.

Першочергові заходи – це конструкторські і технологічні рішення, які виключають генерацію або понижують інтенсивність випромінювання. Спеціальні засоби захисту (екранування джерел випромінювання, фарбування стін у світлі кольори) попереджують розповсюдження і знижують інтенсивність цих випромінювань у виробничих приміщеннях. Очі захищають окулярами або щитками зі склом – світлофільтром. Для захисту шкіри використовують мазі з речовинами – світлофільтрами для цих променів (салол, саліцилово-метиловий ефір та ін.), а також спецодяг з бавовняних тканин і грубововняного сукна. Руки захищають рукавицями.

ВИСНОВКИ

На основі даних щоденного енергоспоживання ми виявили, що перетворення даних у табличний формат і їх відображення досягли зменшення помилки прогнозування на 67% (збільшення на 38% у необроблених відсоткових пунктах точності за межами вибірки) порівняно з нашим базовим рівнем точності, встановлена нашою моделлю прогнозування Prophet.

Ми також випробували простий підхід AutoML для мультикласової класифікації, що призвело до зменшення помилки передбачення на 42% (збільшення на 8% поза навчальною вибіркою) порівняно з нашою моделлю Gradient Boosting і призвело до 81% зменшення помилки прогнозування (збільшення на 46% точності поза навчальною вибіркою) порівняно з нашою моделлю прогнозування Prophet.

Використовуючи підходи, подібні до наведених вище, для моделювання набору даних часових рядів, крім обмеженого підходу, що передбачає лише розгляд методів прогнозування, ми можемо застосувати більш загальні контрольовані методи машинного навчання і досягти кращих результатів для певних типів задач прогнозування.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Готович, В. А., С. В. Марценко, and Т. Л. Щербак. "Створення мобільного апаратно-програмного пристрою моніторингу характеристик якості електроенергії." Збірник наукових праць Інституту проблем моделювання в енергетиці ім. ГЄ Пухова 70 (2014): 98-105.
2. Марценко, Сергій Володимирович. Математичне моделювання та статистичні методи обробки даних вимірювань в задачах моніторингу електронавантаження. Diss. Тернопільський національний технічний університет імені Івана Пулюя, 2011.
3. Готович, Володимир Анатолійович, and В. А. Готович. Математичне моделювання і статистичне оцінювання характеристик штатного режиму електроспоживання організацій. Diss. Тернопільський національний технічний університет імені Івана Пулюя, 2019.
4. Kozlovskiy, Valerii, et al. "Information Technology for Estimating City Gas Consumption During the Year." 2022 International Conference on Smart Information Systems and Technologies (SIST). IEEE, 2022.
5. Lupenko, Serhii, et al. "Approach to gas consumption process forecasting on the basis of a mathematical model in the form of a random cyclic process." Proceedings of the International Conference "Advanced applied energy and information technologies 2021", 2021. TNTU, Zhytomyr «Publishing house "Book-Druk"» LLC, 2021.
6. Lytvynenko, I., et al. "Simulation of gas consumption process based on the mathematical model in the form of cyclic random process considering the scale factors." 1st International Workshop on Information Technologies: Theoretical and Applied Problems, ITAP 2021. 2021.
7. Оробчук, Богдан Ярославович, et al. "Енергоощадна інтелектуальна система керування механічною системою." Матеріали міжнародної науково-технічної конференції „Математичні методи та моделі технічних і економічних

систем “присвячена пам’яті професора Шаблія Олега Миколайовича та 60-ти річчю кафедри теоретичної механіки (2022): 128-130.

8. Бартошевський, Р. В., and Богдан Ярославович Оробчук. "Інтелектуальна система управління та контролю параметрів електричної мережі." Матеріали XI Міжнародної науково-практичної конференції молодих учених та студентів „Актуальні задачі сучасних технологій “ (2022): 82-83.

9. Orobchuk, Bogdan, Oleh Buniak, and Ivan Sysak. "DESIGN OF AN INTELLIGENT SYSTEM TO CONTROL EDUCATIONAL LABORATORY EQUIPMENT BASED ON A HYBRID MINI-POWER PLANT." Eastern-European Journal of Enterprise Technologies 122.9 (2023).

10. Оробчук, Богдан Ярославович, В. Г. Прокопчук, and Р. В. Бартошевський. "Методи зниження втрат електричної енергії в розподільчих мережах." Матеріали XII Міжнародної науково-практичної конференції молодих учених та студентів „Актуальні задачі сучасних технологій “ (2023): 226-227.

11. Executive summary. Global electricity demand rose moderately in 2023 but is set to grow faster through 2026. [Електронний ресурс]. – Режим доступу: <https://www.iea.org/reports/electricity-2024/executive-summary>. 26.03.2024.

12. Hourly Energy Consumption
https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption?select=PJME_hourly.csv [Електронний ресурс]. – Режим доступу: <https://www.iea.org/reports/electricity-2024/executive-summary>. 27.03.2024.

13. Сайт CleanLab. [Електронний ресурс]. – Режим доступу: <https://cleanlab.ai/> 21.04.2024.

14. Стручок, В. С., Стручок, О. С., & Мудра, Д. В. (2017). Навчальний посібник до написання розділу дипломного проекту та дипломної роботи "Безпека в надзвичайних ситуаціях "для студентів всіх спец. денної, заочної (дистанційної) та екстернатної форм навчання.

15. Стручок, В. С. (2022). Техноекологія та цивільна безпека. Частина «Цивільна безпека». Навчальний посібник.

16. Шконда В.В., Кальянов А.В., Давыдов П.Г. Феномен синергетики: наука – общество – образование: Монография / Ред. Шконда В.В. – Донецк: Норд-Пресс, 2009. – 156 с.

17. Жидецький, В. Ц., Джигирей, В. С., & Мельников, О. В. (2000). Основи охорони праці. Львів: Афіша, 350, 132-136.

18. Навакатіян О.О., Кальниш В.В., Стрюков С.М. Охорона праці користувачів комп'ютерних відеодисплейних терміналів. - К.:1997. - 400с.

ДОДАТКИ

Програмний код

```
import pandas as pd
pd.options.mode.chained_assignment = None

data = pd.read_csv('PJME_hourly.csv', parse_dates=['Datetime'],
index_col='Datetime')

# Assuming pjme_data is loaded as before
daily_data = data.resample('D').mean()

# Prepare data for Prophet
daily_data.reset_index(inplace=True)
daily_data.columns = ['ds', 'y']

from sklearn.model_selection import train_test_split

# Reset the datetime
data["Datetime"] = data.index
data = data.reset_index(drop=True)

# Create copy for multiclass data
df = data.copy()

# Convert the datetime column
df['Datetime'] = pd.to_datetime(df['Datetime']) # Adjust the 'datetime' column
name as necessary
df = df.sort_values('Datetime').reset_index(drop=True)

# Obtain day and hour
df['Date'] = pd.to_datetime(df['Datetime']).dt.floor('D')
df['Hour'] = pd.to_datetime(df['Datetime']).dt.hour

# Create multi-index feature df to compute time series features on
features = df.set_index(['Date', 'Hour'])
features.drop("Datetime", inplace=True, axis=1)

# Split the data into training and testing sets, respecting the temporal order
X_train, X_test, y_train, y_test = train_test_split(features, features["PJME_MW"],
test_size=0.2, shuffle=False)

# Get group lengths
train_lengths = X_train.groupby(level=0).size()
test_lengths = X_test.groupby(level=0).size()

# Obtain common length value for train/test data
train_common_length = train_lengths.mode().iloc[0]
test_common_length = test_lengths.mode().iloc[0]

# Filter train/test data to groups with same common length for featurizer
X_train = X_train.groupby(level=0).filter(lambda x: len(x) == train_common_length)
X_test = X_test.groupby(level=0).filter(lambda x: len(x) == test_common_length)

# Create quartiles based on training data to avoid leakage
quartiles = [X_train['PJME_MW'].quantile(q) for q in [0.25, 0.50, 0.75]]
```

```
# Cutoff date at 2015-04-09
cutoff_index = int(len(daily_data) * 0.8)

# Use 80% of data for training set and 20% for test set
train_df = daily_data.iloc[:cutoff_index]
test_df = daily_data.iloc[cutoff_index:]
```

```
print("Training Set Shape:", train_df.shape)
print("Testing Set Shape:", test_df.shape)
Training Set Shape: (4847, 2)
Testing Set Shape: (1212, 2)
train_df.tail()
```

	ds	y
4842	2015-04-05	24577.500000
4843	2015-04-06	26996.666667
4844	2015-04-07	27177.833333
4845	2015-04-08	29136.041667
4846	2015-04-09	30535.291667

```
test_df.head()
```

	ds	y
4847	2015-04-10	29190.166667
4848	2015-04-11	24774.291667
4849	2015-04-12	24407.625000
4850	2015-04-13	26825.333333
4851	2015-04-14	26952.125000

```
import numpy as np
from prophet import Prophet
from sklearn.metrics import accuracy_score
```

```
# Initialize model and train it on training data
model = Prophet()
model.fit(train_df)
```

```
# Create a dataframe for future predictions covering the test period
future = model.make_future_dataframe(periods=len(test_df), freq='D')
forecast = model.predict(future)
```

```
# Categorize forecasted daily values into quartiles based on the thresholds
forecast['quartile'] = pd.cut(forecast['yhat'], bins = [-np.inf] + list(quartiles)
+ [np.inf], labels=[1, 2, 3, 4])
```

```
# Extract the forecasted quartiles for the test period
forecasted_quartiles = forecast.iloc[-len(test_df):]['quartile'].astype(int)
```

```
# Categorize actual daily values in the test set into quartiles
test_df['quartile'] = pd.cut(test_df['y'], bins=[-np.inf] + list(quartiles) +
[ np.inf], labels=[1, 2, 3, 4])
actual_test_quartiles = test_df['quartile'].astype(int)
```

```
# Calculate the evaluation metrics
```

```

prophet_accuracy = accuracy_score(actual_test_quartiles, forecasted_quartiles)

# Print the evaluation metrics
print(f'Accuracy: {prophet_accuracy:.4f}')
/Users/mturk/mturk-work/cleanlab-tools/time-series-automl-env/lib/python3.11/site-
packages/tqdm/auto.py:21: TqdmWarning: IProgress not found. Please update jupyter
and ipywidgets. See https://ipywidgets.readthedocs.io/en/stable/user_install.html
  from .autonotebook import tqdm as notebook_tqdm
Importing plotly failed. Interactive plots will not work.
13:04:35 - cmdstanpy - INFO - Chain [1] start processing
13:04:35 - cmdstanpy - INFO - Chain [1] done processing
Accuracy: 0.4249
# For illustrative purposes we show the quartiles for training data
train_df['quartile'] = pd.cut(train_df['y'], bins = [-np.inf] + list(quartiles) +
[np.inf], labels=[1, 2, 3, 4])
train_df['quartile'].head()
0      2
1      3
2      3
3      3
4      2
Name: quartile, dtype: category
Categories (4, int64): [1 < 2 < 3 < 4]

import tsfel
from sktime.transformations.panel.tsfresh import TSFreshFeatureExtractor

# Define tsfresh feature extractor
tsfresh_trafo = TSFreshFeatureExtractor(default_fc_parameters="minimal")

# Transform the training data using the feature extractor
X_train_transformed = tsfresh_trafo.fit_transform(X_train)

# Transform the test data using the same feature extractor
X_test_transformed = tsfresh_trafo.transform(X_test)

# Retrieves a pre-defined feature configuration file to extract all available
features
cfg = tsfel.get_features_by_domain()

# Function to compute tsfel features per day
def compute_features(group):
    # TSFEL expects a DataFrame with the data in columns, so we transpose the
input group
    features = tsfel.time_series_features_extractor(cfg, group, fs=1, verbose=0)
    return features

# Group by the 'day' level of the index and apply the feature computation
train_features_per_day =
X_train.groupby(level='Date').apply(compute_features).reset_index(drop=True)
test_features_per_day =
X_test.groupby(level='Date').apply(compute_features).reset_index(drop=True)

# Combine each featurization into a set of combined features for our train/test
data
train_combined_df = pd.concat([X_train_transformed, train_features_per_day],
axis=1)
test_combined_df = pd.concat([X_test_transformed, test_features_per_day], axis=1)

# Filter out features that are highly correlated with our target variable

```

```

column_of_interest = "PJME_MW_mean"
train_corr_matrix = train_combined_df.corr()
train_corr_with_interest = train_corr_matrix[column_of_interest]
null_corrs = pd.Series(train_corr_with_interest.isnull())
false_features = null_corrs[null_corrs].index.tolist()

columns_to_exclude =
list(set(train_corr_with_interest[abs(train_corr_with_interest) >
0.8].index.tolist() + false_features))
columns_to_exclude.remove(column_of_interest)

# Filtered DataFrame excluding columns with high correlation to the column of
interest
X_train_transformed = train_combined_df.drop(columns=columns_to_exclude)
X_test_transformed = test_combined_df.drop(columns=columns_to_exclude)
Feature Extraction: 100%| ████████████████████████████████████████████████ |
4817/4817 [00:00<00:00, 7909.95it/s]
Feature Extraction: 100%| ████████████████████████████████████████████████ |
1205/1205 [00:00<00:00, 9343.20it/s]
# Define a function to classify each value into a quartile
def classify_into_quartile(value):
    if value < quartiles[0]:
        return 1
    elif value < quartiles[1]:
        return 2
    elif value < quartiles[2]:
        return 3
    else:
        return 4

y_train = X_train_transformed["PJME_MW_mean"].rename("daily_energy_level")
X_train_transformed.drop("PJME_MW_mean", inplace=True, axis=1)

y_test = X_test_transformed["PJME_MW_mean"].rename("daily_energy_level")
X_test_transformed.drop("PJME_MW_mean", inplace=True, axis=1)

energy_levels_train = y_train.apply(classify_into_quartile)
energy_levels_test = y_test.apply(classify_into_quartile)

from sklearn.ensemble import GradientBoostingClassifier

gbc = GradientBoostingClassifier(
    n_estimators=150,
    learning_rate=0.1,
    max_depth=4,
    min_samples_leaf=20,
    max_features='sqrt',
    subsample=0.8,
    random_state=42
)

gbc.fit(X_train_transformed, energy_levels_train)

y_pred_gbc = gbc.predict(X_test_transformed)
gbc_accuracy = accuracy_score(energy_levels_test, y_pred_gbc)
print(f'Accuracy: {gbc_accuracy:.4f}')
Accuracy: 0.8075

```

Here's all the code needed to train and deploy an AutoML supervised classifier:

```
from cleanlab_studio import Studio

studio = Studio()
studio.create_project(
    dataset_id=energy_forecasting_dataset,
    project_name="ENERGY-LEVEL-FORECASTING",
    modality="tabular",
    task_type="multi-class",
    model_type="regular",
    label_column="daily_energy_level",
)

model = studio.get_model(energy_forecasting_model)
y_pred_automl = model.predict(test_data, return_pred_proba=True)
y_pred_automl_cleanlab = pd.read_csv("quartile-multiclass-pjme-testing-
data_pred_probs.csv")
y_pred_automl_cleanlab = y_pred_automl_cleanlab["Suggested Label"]
automl_accuracy = accuracy_score(energy_levels_test, y_pred_automl_cleanlab)
print(f'Accuracy: {automl_accuracy:.4f}')
Accuracy: 0.8880
# Calculate error rates
error_rate_prophet = 1 - prophet_accuracy
error_rate_gbc = 1 - gbc_accuracy
error_rate_automl = 1 - automl_accuracy

print(f"Prophet error rate: {error_rate_prophet}")
print(f"GBC error rate: {error_rate_gbc}")
print(f"AutoML error rate: {error_rate_automl}")

# Calculate reduction in prediction error
error_reduction_gbc = error_rate_prophet - error_rate_gbc
error_reduction_automl_to_prophet = error_rate_prophet - error_rate_automl
error_reduction_automl_to_gbc = error_rate_gbc - error_rate_automl

# Convert error reduction to a percentage of improvement from one model to another
percentage_improvement_gbc = round((error_reduction_gbc / error_rate_prophet) *
100, 1)
percentage_improvement_automl_to_prophet =
round((error_reduction_automl_to_prophet / error_rate_prophet) * 100, 1)
percentage_improvement_automl_to_gbc = round((error_reduction_automl_to_gbc /
error_rate_gbc) * 100, 1)

print(f"GBC compared to Prophet resulted in a {percentage_improvement_gbc}%
reduction in prediction error.")
print(f"AutoML compared to Prophet resulted in a
{percentage_improvement_automl_to_prophet}% reduction in prediction error.")
print(f"AutoML compared to GBC resulted in a
{percentage_improvement_automl_to_gbc}% reduction in prediction error.")
Prophet error rate: 0.5750825082508251
GBC error rate: 0.1925311203319502
AutoML error rate: 0.11203319502074693
GBC compared to Prophet resulted in a 66.5% reduction in prediction error.
AutoML compared to Prophet resulted in a 80.5% reduction in prediction error.
AutoML compared to GBC resulted in a 41.8% reduction in prediction error.
```