

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
Сумський державний університет  
Факультет електроніки та інформаційних технологій  
Кафедра інформаційних технологій

«До захисту допущено»

В.о. завідувача кафедри

\_\_\_\_\_ Світлана ВАЩЕНКО

\_\_\_\_\_ 2024 р.

**КВАЛІФІКАЦІЙНА РОБОТА**  
**на здобуття освітнього ступеня магістр**

зі спеціальності 122 «Комп'ютерні науки»,

освітньо-наукової програми «Інформаційні технології проектування»

на тему: **«Моделі інформаційної технології прогнозування електроспоживання підприємства в умовах невизначеності»**

Здобувача групи ІТ.м-21н Доценка Олександра Романовича  
(шифр групи) (прізвище, ім'я, по батькові)

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

\_\_\_\_\_  
(підпис)

Олександр Доценка  
(Ім'я та ПРІЗВИЩЕ здобувача)

Керівник професор кафедри ІТ, д.т.н., доцент Сергій ТИМЧУК  
(посада, науковий ступінь, вчене звання, Ім'я та ПРІЗВИЩЕ)

\_\_\_\_\_  
(підпис)

Сумський державний університет  
Факультет електроніки та інформаційних технологій  
Кафедра інформаційних технологій  
Спеціальність 122 «Комп'ютерні науки»  
Освітньо-наукова програма «Інформаційні технології проектування»

**ЗАТВЕРДЖУЮ**

В.о. завідувача кафедри ІТ

Світлана ВАЩЕНКО

« \_\_\_\_\_ » \_\_\_\_\_ 2024 р.

## **ЗАВДАННЯ**

**на кваліфікаційну роботу магістра студентіві**

Доценко Олександр Романович

(прізвище, ім'я, по батькові)

**1 Тема кваліфікаційної роботи** Моделі інформаційної технології прогнозування електроспоживання підприємства в умовах невизначеності

затверджена наказом по університету від «01» лютого 2024 р. № 0096-VI

**2 Термін здачі студентом кваліфікаційної роботи** «10» травня 2024 р.

**3 Вхідні дані до кваліфікаційної роботи** перелік вимог до моделі інформаційної технології, файли історичних замірів даних

**4 Зміст розрахунково-пояснювальної записки (перелік питань, що їх належить розробити)** Аналіз предметної області; Актуальність проблеми; Опис моделей і методів, які застосовуються в умовах невизначеності; Обґрунтування вибору конкретної моделі чи методу; Розробка та тестування програми; Презентація.

**5 Перелік графічного матеріалу (з точним зазначенням обов'язкових слайдів презентації)** Об'єкт, предмет, мета та гіпотеза; Актуальність дослідження; Функціональні вимоги до моделі; Вибір методів прогнозування; Процес розробки: Програмна реалізація створеної моделі; Тестування; Демонстрація роботи; Висновки;

## 6. Консультанти випускної роботи із зазначенням розділів, що їх стосуються:

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

Дата видачі завдання \_\_\_\_\_.

Керівник \_\_\_\_\_  
(підпис)

Завдання прийняв до виконання \_\_\_\_\_  
(підпис)

### КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Термін виконання етапів роботи	Примітка
	Аналіз предметної області	30.11.2023-22.02.2024	
	Визначення вимог	23.01.2024-16.02.2024	
	Аналіз і вибір методу реалізації	17.02.2024-15.03.2024	
	Проектування архітектури програми	16.03.2024-10.04.2024	
	Реалізація моделі прогнозування	11.04.2024-25.04.2024	
	Тестування та виправлення помилок реалізації	26.04.2024-30.04.2024	
	Завершення проекту	01.05.2024-10.05.2024	
	Презентація	22.05.2024	

Магістрант \_\_\_\_\_

Доценко О.Р.

Керівник роботи \_\_\_\_\_

професор кафедри ІТ, д.т.н.,  
доцент Тимчук С.О.

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

- АР – метод авторегресії
- ВДЕ – відновлювана енергія
- ЕЕ – електроенергія
- ЕЕС – електроенергетична система
- ЕМ – економетричні методи
- ЕС – експертна система
- ІТ – інформаційні технології
- КС – метод ковзного середнього
- МН – машинне навчання
- НРА – нечіткий регресійний аналіз
- ОРЕ – оптовий ринок електроенергії
- ОУ – об'єкт управління
- ПЗ – програмне забезпечення
- СМ – статистичні методи
- ШІ – штучний інтелект

## АНОТАЦІЯ

Кваліфікаційна робота магістра на тему «Моделі інформаційної технології прогнозування електроспоживання підприємства в умовах невизначеності» є актуальною через необхідність ефективно використовувати енергію. Робота складається зі вступу, чотирьох розділів, висновку, списку використаних джерел на 44 найменувань та додатку. Весь файл має 66 сторінок, де 46 сторінки є основний текст, 6 сторінок використаних джерел, 5 сторінок додатків.

Метою дослідження є розробка та впровадження моделі інформаційної технології для прогнозування електроспоживання підприємства в умовах невизначеності. Робота передбачає аналіз різних методів прогнозування та їх ефективності в конкретних умовах функціонування установи. Важливим аспектом є аналіз інформаційних потреб компанії та розробка моделі, яка враховує непередбачувані зміни у споживанні електроенергії. Також, ретельне спостереження потреб та особливостей енергоспоживання, щоб забезпечити максимальну адаптивність і точність.

Одним із ключових результатів виконання цієї програми є тестування прогнозної моделі (лінійної регресії з оптимізацією градієнтного спуску), на реальних даних. Ця програма має вирішальне значення для практичного застосування, особливо в системах підтримки прийняття рішень, оскільки вона допомагає підвищити ефективність управління шляхом надання точних прогнозів.

Дана розробка може бути інтегрована у веб-додатку для взаємодії з користувачем за допомогою фреймворків Flask або Django та розгорнута на хмарних платформах щоб прогнозувати не тільки локальні енергомережі.

Ключові слова: моделі прогнозування, статистичні методи, машинне навчання, прогнозування електроенергії, інформаційні технології, лінійна регресія, мова програмування Python, регресійне прогнозування в Python.

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ.....	4
ВСТУП.....	7
1. АНАЛІТИЧНИЙ ОГЛЯД .....	9
1.1 Прогнозування даних енергоспоживання, генерації і вартості енергії .....	9
1.2 Значення проблеми прогнозування в контексті ринку електроенергії. ....	12
1.3 Використання штучного інтелекту в прогнозуванні електроспоживання ....	15
2. ВИБІР МЕТОДІВ ПРОГНОЗУВАННЯ .....	20
2.1 Існуючі способи прогнозування електроенергії .....	20
Метод авторегресії та ковзного середнього. ....	20
Експоненційне згладжування. ....	22
Метод найближчого сусіда. ....	25
2.2 Нечіткий регресійний аналіз .....	28
2.3 Інформаційні технології прогнозування електроенергії .....	32
3. ВИБІР ОПТИМАЛЬНОГО МЕТОДУ ПРОГНОЗУВАННЯ.....	36
3.1 Порівняння та аналіз моделей прогнозування .....	36
Статистичні моделі .....	36
Машинне навчання .....	39
Економетричні методи.....	41
3.2 Обраний метод.....	43
3.3 Мова програмування Python .....	44
4. СТВОРЕННЯ ТА ОПИСАННЯ ПРОГРАМИ .....	48
ВИСНОВОК .....	56
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....	57
ДОДАТОК А .....	63
ДОДАТОК Б.....	67

## ВСТУП

На сьогоднішній момент значення електрики в усіх сферах життєдіяльності сучасного суспільства просто колосальне, а в умовах постійної зміни технологічного прогресу та високої конкуренції на ринку, ефективне управління електроспоживанням підприємств стає ключовим фактором для забезпечення їхньої стабільності та конкурентоспроможності. Якісного та безперебійного підключення потребують заклади освіти та охорони здоров'я, промислові та сільськогосподарські підприємства, науково-дослідницькі та фінансово-економічні центри, урядові та оборонні структури. Прогнозування електроспоживання в умовах невизначеності є важливою складовою цього процесу, оскільки дозволяє уникнути несподіваних витрат та оптимізувати використання матеріалів.

Також з кожним роком інформаційні технології захоплюють все більше і більше сфер людського життя. Вони стали повсякденним помічником, роблячи наш світ набагато пов'язаним та інноваційним. ІТ – широкий термін для позначення комп'ютерів, ПЗ, мереж та інших технологій для зберігання, передачі та обробки інформації. Маючи великий потенціал та постійний вплив на нас в житті, бізнесі, освіті не дивно, що інновації заходять в енергетичну сферу і допомагають підприємствам розвиватися.

**Актуальність:** З початком широкомасштабного наступу РФ на Україну, енергетична система зазнає все більше і більше руйнувань. Ці терористичні дії призвели до руйнувань понад 300 об'єктів теплоенергетики і третину енергетичної інфраструктури. З цими проблемами не зіштовхувалася ще жодна країна у сучасному світі. Тому сучасні методи та моделі допоможуть враховувати різноманітні фактори, що впливають на споживання ресурсу. Зокрема, описуються методи статистичного аналізу, математичні моделі, а також використання штучного інтелекту для покращення точності прогнозів. Врахування цих факторів дозволяє побудувати більш точні та надійні прогнози, що є критично важливим для оптимізації виробничих процесів та планування можливостей.

**Об'єкт дослідження:** процес розробки та вдосконалення моделі прогнозування, котрий дозволяє передбачити обсяг електроенергії та в подальшому використовувати підприємством в майбутньому.

**Предмет дослідження:** аналіз різних підходів та методів для створення моделі, яка дозволить передбачувати споживання ЕЕ.

**Гіпотеза:** використання методів нечіткої математики сприятиме покращенню точності та надійності прогнозів електроспоживання, дозволяючи підприємствам більш ефективно планувати свою роботу та ресурси.

**Мета:** розглянути і проаналізувати моделі інформаційної технології прогнозування ЕЕ та застосувати їх в умовах невизначеності.

**Наукова новизна:** набув подальшого розвитку метод інтеграції адаптивних прогнозних моделей в інформаційну технологію, що дозволило оперативно враховувати змінні умови та динамічність даних електроспоживання.

**Практичне значення:** інформаційна технологія, що включає модель передбачення електроспоживання в умовах невизначеності і її практичне застосування дозволить покращити ефективність управління електроспоживання для підвищення конкурентоспроможності на ринку.

Дослідження в цьому напрямку має принести новітні результати, розвиваючи підходи до моделювання споживання в змінливих умовах. Отримані дані матимуть не лише академічне значення, але й практичне застосування для підприємств, що мають амбіції зменшити залежність від зовнішніх факторів та підвищити енергоефективність своєї діяльності. У цьому контексті відкривається можливість переосмислення стратегій управління енергетичними ресурсами та створення основи для подальших досліджень у сфері інформаційних технологій та прогнозування енергоспоживання.



## 1. АНАЛІТИЧНИЙ ОГЛЯД

### 1.1 Прогнозування даних енергоспоживання, генерації і вартості енергії

Еволюція економіки залежить від планування та спостереження її зростання, а також від розробки систем управління фінансами країни. Прогнозування енергетичних тенденцій відіграє важливу роль у контролі ефективності використання енергії та формуванні стратегій і енергетичної політики. Це також є ключовим елементом у формулюванні стратегій і енергетичних законів.

Прорахунок значень на основі минулих даних в енергетичному секторі є життєво важливою ціллю для фінансового планування, нагляду за рівнями виробництва електрики та підтримки контролю. Антиципація цін має ключові значення в цьому сегменті, оскільки вони значною мірою впливають на рішення щодо довгострокових інвестицій у нові генеруючі потужності виробничої діяльності. [1] Усі учасники ринку повинні знати про майбутні цінники на електроенергію, оскільки їх прибутковість хвилює і виробників компаній ЕЕ, і інвесторів, і великих промислових споживачів тощо.

Безперервне функціонування економіки залежить від прогнозування та стратегічного планування її розвитку. Енергетичне покладання є першочерговим завданням у цьому контексті враховуючи, що більше половини паливних ресурсів спрямовується на виробництво електроенергії, яка робить це завдання прорахунку розвитку енергетики особливо актуальним [2]. Ці прогнози є важливими не лише з точки зору регіональної економіки, а й для найбільших енергетичних компаній по всьому світу.

Коли мова заходить про енергетичне передбачення значень часових рядів, використання історичних даних має вирішальне значення для фінансового планування, управління рівнями виробництва енергії та підтримки контролю. Їхні обов'язки класифікуються на основі відхилення, часовими рамками горизонту, для якого значення часового ряду має бути визначено [3]:

- Короткострокове прогнозування

- Середньострокове прогнозування
- Довгострокове прогнозування

На рисунку 1.1 наведені дані типу прогнозів, їхні методи прогнозування, кількість часу тощо. Вона краще та ширше надасть інформацію та наглядно продемонструє кожну із генерацій.

	Тип прогнозу	Горизонт часу	Типове застосування	Метод прогнозування
Генерація	внутрішньогодинний	5-60 хвилин	Регулювання, диспетчеризація в режимі реального часу, ринковий кліринг і взаємозаліки	Статистичні: персистенція
	короткостроковий	На 1-6 години наперед	Розробка графіків, відстежування навантаження споживачів, управління перевантаженнями	Суміш статистичних і чисельних методів прогнозування погоди (ЧПП)
	середньостроковий	На день наперед	Розробка графіків, резервування потужностей і завчасний пуск резервних потужностей, прогноз перетікань, управління перевантаженнями	ЧПП з корекцією систематичної похибки
	довгостроковий	Тижневий, сезонний, річний або більше за 1 рік	Ресурсне планування, аналіз наслідків аварійних порушень режиму, планування обслуговування, оперативне управління	Кліматологічні прогнози + ЧПП
	Прогноз генерації	постійний	Ситуаційний контроль, процедура скорочення об'єму угод з генерації (через пропускну здатність мереж)	ЧПП + статистичні методи
Підтримка ухвалення рішення	Прогноз споживання (навантаження)	На день наперед, на годину наперед, внутрішньогодинний	Розробка графіків, економічна диспетчеризація, управління перевантаженнями управління електроспоживання споживачів	Статистичні

Рисунок 1.1 – Наглядна таблиця типів прогнозів

В даний момент існує понад 100 категорій моделей прогнозування, деякі з яких належать до окремих методів. Вони поділяються на дві групи: інтуїтивні та формалізовані. Інтуїтивне передбачення використовується, коли об'єкт, над яким

робляться досліди, занадто простий або надто складний, що унеможливує врахування впливу зовнішніх факторів. Такі методи відображають лише індивідуальні судження спеціалістів щодо потенційного розвитку процесів і не сприяють створенню прогнозу моделі. Інтуїтивно зрозуміла методика зазвичай використовуються для аналізу процесуальних дій, котрі неможливо формалізувати математично або для яких складно розробити відповідну прогностичну систему. До них входять такі аспекти: вибірковий прогноз, історичні аналогії, експертні оцінки [4]. Формалізовані методики можна далі поділити на статистичні та структурні прогностичні аналізи. Статистичні включають аналітичну специфікацію, як зовнішніх факторів, так і функціональні залежності між майбутніми та фактичними значеннями часового ряду. Ці моделі представлені такими групами:

**Регресійні моделі:** використовуються для прогнозування електричного навантаження енергосистем, виробничих приміщень, будівель.

**Авторегресійні моделі:** цей підхід розробився для короткострокового передбачення споживання електроенергії в робочій ітерації з інтервалом попередження 5-10 хвилин в роботі.

**Експоненціальні моделі згладжування,** спочатку розроблені Холтом і Брауном у 20 столітті, експлуатуються в роботах для річного прогнозу споживання ЕЕ промисловими підприємствами та енергетичними системами. Завдання передбачування річного обсягу виробництва відновлюваної енергії (ВДЕ) також вирішується за допомогою техніки експоненціального згладжування в різних дослідженнях [3].

Структурні моделі, в котрих задається функціональна залежність між майбутніми і теперішніми значеннями часового ряду та зовнішніми факторами, охоплюють декілька груп. До них належать моделі на основі ланцюгів Маркова, моделі нейронних мереж та моделі на основі дерев класифікації та регресії. В статті [5], модель, що використовує нейротехнології, використовується для оцінки потужності та втрат електроенергії в системах електромереж. Інше дослідження [6] використовує алгоритми нейронної мережі для короткострокового прогнозування навантаження центральної електроенергетичної системи країни. Деякі автори [7]

розглядають як короткострокове, так і довгострокове прогнозування навантажень енергосистеми в цілому.

Критичними потребами до моделей прогнозування є бажано висока точність передбачення та нескладний алгоритм, що забезпечує мінімальний об'єм рішення та кількість системної пам'яті. Крім того, ці моделі повинні функціонувати в умовах невизначеності та недостатньої інформації, зберігаючи стабільність управління.

## **1.2 Значення проблеми прогнозування в контексті ринку електроенергії.**

Актуальність прогнозування на ринку електроенергії полягає в значущості прогнозування режимних параметрів і техніко-економічних показників, які є визначальними як при плануванні, так і при діючих режимах електроенергетичної системи (ЕЕС) [3]. Створення умов для вільної конкуренції є складовою реформування галузі, спрямованої на баланс інтересів виробників і споживачів. Помилки прогнозування та неефективне споживання ресурсів стають дорогими проблемами. Дослідження, проведені для енергокомпаній окремої країни, показали, що навіть незначне покращення якості прогнозування може значно знизити витрати на відхилення від плану постачання електроенергії.

Мінімізація помилок і розробка ПЗ, яке може працювати з обмеженою кількістю ретроспективної інформації, є актуальним завданням. Під час переходу до ринкової економіки електроенергія перетворюється на товар, який характеризується не лише кількістю, але й вартістю. Формування конкурентоспроможної ціни є необхідною умовою залучення інвестицій у розвиток енергетики. Оптовий ринок електроенергії та електроенергії (ОРЕ) охоплює систему договірних відносин між учасниками, які поєднують свої технологічні процеси виробництва, передачі, розподілу та споживання електроенергії. Його можна розділити на чотири основні сегменти: ринок на добу наперед (спот), ринок потужності, балансуєчий ринок і сегмент довгострокових контрактів (рис. 1.2) [8].

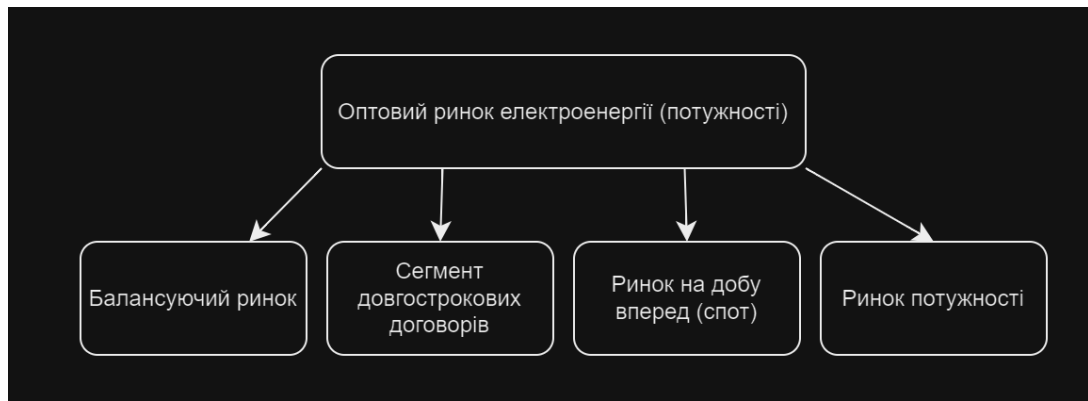


Рисунок 1.2 – Основні сегменти оптового ринку електроенергії

**Сегмент довгострокових договорів** становить основу ОРЕ, де більша частина електроенергії "розпродається". В цьому сегменті обсяги та ціни в прямих договорах між виробником і споживачем встановлені державою на певний період, тож всі 100% з обсягу учасників схеми закріплені в таких договорах [9]. У майбутньому цей сегмент буде щорічно скорочуватись за рахунок збільшення обсягів ресурсу, що поставляється за вільними довгостроковими договорами і на самому маркеті.

**Балансуючий ринок** є "онлайн" платформою, де розторгуються заявки на поточне виробництво або споживання на наступну годину вперед до фактичного виробництва/споживання. Тут торгуються обсяги відхилень фактичного виробництва/споживання від запланованого після результатів торгів на ОРЕ.

**Ринок на добу вперед (спот)** - це місце торгівлі електроенергією на наступний день. Тут формується ціна ресурсу в залежності від обмежень по передачі енергії, мережевих втрат та інших параметрів. Спочатку тут продається енергія в обсязі відхилень від параметрів, що зафіксовані в регульованих договорах.

**Ринок потужності** - місце аукціонів на постачання/покупку потужності енергії на рік, а згодом на 3 роки вперед. Він дозволяє електростанціям отримувати кошти, які компенсують умовно-постійні витрати, пов'язані з підтримкою у робочому стані генеруючого обладнання [8].

На цьому маркеті постачальниками є генеруючі компанії та імпортери, а покупцями виступають споживачі, збутові компанії і експортери електроенергії [10]. Перехід на довгострокові двосторонні відносини в умовах лібералізації ринку

забезпечує прогнозованість вартості струму у середньостроковій та довгостроковій перспективі, що є важливою умовою інвестиційної привабливості сектору.

Ринок "на добу вперед" базується на конкурентному відборі цінових замовлень від постачальників та покупців за добу від реальної поставки електроенергії, з визначенням цін та обсягів поставки на кожну годину протягом дня [11]. У випадку відхилення від запланованих обсягів поставки, учасники здійснюють купівлю або продаж на балансуєчому ринку. Він замінює собою попередню модель вільної торгівлі, де учасники подавали заявки лише на частку обсягів виробництва або споживання. Результати аукціону цінових замовлень використовуються для планування режимів виробництва та споживання електроенергії.

З проблемою організації прогнозування споживання та обґрунтування заявок за ціною та обсягом на ринку «на добу наперед» стикаються оперативно-диспетчерські служби промислових підприємств. Неправильне планування може призвести до значної кількості штрафів за відхилення між фактичним і запланованим споживанням. Стійкі суб'єкти повинні надавати пріоритет розробці ефективних моделей прогнозування, оскільки сума штрафів прямо пропорційна середньорічній похибці прогнозування споживання електроенергії [12].

Для участі в конкурентному ринку необхідна наявність автоматизованої системи комерційного обліку, що відповідає технічним вимогам ОРЕ. Компанії, які працюють на цьому ринку, зобов'язані здійснювати погодинне планування споживання за два дні до фактичного використання, підкреслюючи важливість короткострокового прогнозування споживання електроенергії та необхідність для підприємств ретельно організувати робочі процеси [10].

Як член Енергетичного Співтовариства, Україна віддана впровадженню принципів третього енергетичного пакету ЄС. Це включає, зокрема, відокремлення генерації та постачання, підвищення прозорості ринків електроенергії, посилення захисту споживачів, створення більш ефективних і незалежних національних регуляторів і покращення транскордонного співробітництва та інвестицій. У 2013 році Верховна Рада України почала відкривати оптовий ринок електроенергії та посилювати конкуренцію. На шляху до інтеграції з європейською мережею країна

почала створювати конкурентні роздрібні та ОРЕ. У 2017 році ВР ухвалила Закон про модель електроенергії, який скасував схему єдиного покупця та запровадив проринкові елементи, які сприяли конкуренції [13]. Тепер вона має такий вигляд (рис. 1.3).

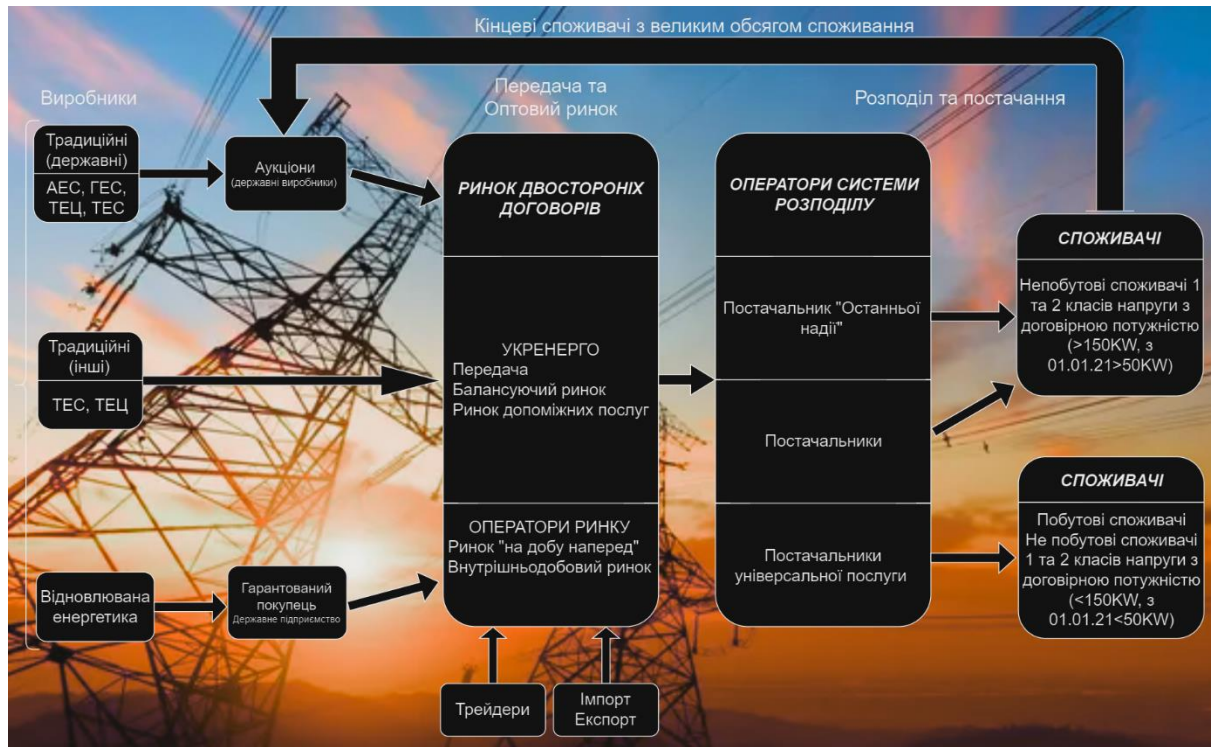


Рисунок 1.3 – Нова модель ринку електричної енергії

### 1.3 Використання штучного інтелекту в прогнозуванні електроспоживання

Штучний інтелект (AI) — це сучасна технологія, яка широко використовується, особливо в сфері цифрових досягнень. У ньому в основному вживаються алгоритми для керування та обробки величезних обсягів даних, покращуючи таким чином різні аспекти функціонування суспільства. Незважаючи на широке поширення, застосування цієї технології залишається значною мірою нестандартизованим. Крім того, остаточне тлумачення терміну «штучний інтелект» ще не встановлено, оскільки дослідження продовжують вивчати його потенційні можливості застосування в різних

секторах [14]. Увага буде зосереджена переважно на його потенційному використанні в енергетичному секторі.

Кабінет міністрів схвалив визначення штучного інтелекту у законодавстві України, зокрема, у схваленій Концепції розвитку штучного інтелекту в Україні (грудень 2020 року) [15], де ШІ це: «організована сукупність інформаційних технологій, із застосуванням якої можливо виконувати складні комплексні завдання шляхом використання системи наукових методів досліджень і алгоритмів обробки інформації, отриманої або самостійно створеної під час роботи, а також створювати та використовувати власні бази знань, моделі прийняття рішень, алгоритми роботи з інформацією та визначати способи досягнення поставлених завдань». Ця технологія може похвалитися багатофункціональністю, що дозволяє створювати численні варіанти використання або шаблони, які можуть одночасно співіснувати в одній системі, кожен з яких призначений для різних завдань. Однак наш інтерес полягає в конкретних випадках використання:

**Цілеспрямовані алгоритми:** вони зосереджені на досягненні цілей за допомогою машинного навчання та інших когнітивних обчислювальних підходів. Використовуючи ітеративний процес навчання методом проб і помилок, ці алгоритми прагнуть знайти оптимальне рішення проблеми [16]. Процес прийняття рішень зазвичай оцінюється за попередньо визначеною функцією витрат. Загальні програми включають оптимізацію ресурсів/логістики, ітераційне вирішення проблем, аукціони та моделювання сценаріїв.

**Алгоритми виявлення шаблонів і аномалій:** ці алгоритми спеціалізуються на розпізнаванні шаблонів у даних і порівнянні їх із усталеними шаблонами. Це порівняння допомагає визначити, чи дані узгоджуються з існуючим шаблоном, чи демонструють відхилення [17]. Машинне навчання відіграє вирішальну роль у ідентифікації таких наборів даних. Додатки включають людські помилки, виявлення шахрайства тощо.

**Алгоритми прогнозування та підтримки прийняття рішень:** ці алгоритми обробляють інформацію, включаючи описову аналітику, прогнозну аналітику та проєктивну аналітику. Вони прогнозують майбутні значення даних, передбачають



поведінку населення, визначають оптимальний підхід і оптимізують операції. Прогнозування зазвичай використовує історичні та існуючі поведінкові дані для прогнозування майбутніх результатів, тим самим допомагаючи прийняттю рішень.

Ці сценарії використання можна застосовувати окремо або комбінувати для вирішення конкретних завдань у автоматизованих системах III. Унікальні аспекти послуг енергопостачання, такі як необхідність обробки величезних обсягів даних у режимі реального часу, пов'язаних із прогнозом погоди, енергоспоживанням і обсягами виробництва, станом обладнання та роботою ліній електропередач, відкривають широкі можливості для застосування штучного інтелекту в секторі енергетики. [18] Наприклад: прогнозування погодних умов може передбачати зміни в попиті споживачів на енергію, таким чином дозволяючи краще планувати генеруючі потужності та підвищувати ефективність існуючих установок. Він також може стабілізувати системи передачі енергії, виявляючи похибки у моделях виробництва та споживання. Може розробляти рішення в режимі реального часу для усунення цих аномалій, наприклад підключення або відключення джерел енергії або додаткового обладнання, забезпечуючи стабільність і надійність системи.

Підвищена доступність даних за допомогою цієї технології дозволяє покращити прогнозування режимів роботи системи розподілу енергії: планування технічного обслуговування в електричних мережах можна оптимізувати, використовуючи інформацію про проектний термін служби, непроекtnі режими роботи та рівні зносу окремих елементів, що значно скорочує аварійні ситуації в мережі та дорогі простої. Для споживачів енергії штучний інтелект дозволяє оптимізувати споживання енергії шляхом прогнозування цін на енергію в різні часові проміжки та управління попитом з боку споживачів за допомогою розумного обладнання [19]. Режим роботи приладу можна регулювати залежно від звичок клієнтів (наприклад, автоматичний час увімкнення кондиціонера або освітлення), фактичних потреб у споживанні (наявність споживача в кімнаті) та витрат на електроенергію в певні періоди, щоб перенести графік споживання на час, коли енергія дешевша [14].

Особливо в нинішньому контексті повномасштабного російського військового вторгнення в Україну та навмисного знищення всього енергетичного сектору точне

прогнозування стало критичним питанням для нашого суспільства. Важливе значення для нашої країни має енергетична галузь, яка становить 8% ВВП, займає перше місце серед платників податків до державного бюджету та забезпечує роботою 450 тис. осіб, що становить приблизно 3% населення [2]. Незважаючи на виклики, сектор зберігає значний потенціал для подальшого зростання та залишається привабливою перспективою як для національних, так і для міжнародних інвесторів.

Тому з'являються такі проекти, як українська компанія A-Gnostics, котра пропонує послугу промислового штучного інтелекту, яка спеціалізується на виявленні аномалій і прогнозуванні несправностей обладнання. Ця послуга налаштована для багатофакторних процесів і даних часових рядів, отриманих від промислового обладнання, що дає змогу автоматично й точно визначати нормальні, несправні та передвідмовні умови. Основною метою є використання машинного навчання та штучного інтелекту для завчасного прогнозування та запобігання порушень обладнання. Серед їхніх інноваційних пропозицій є Di-agnostics, яку часто називають «Shazam для промислового обладнання». У цьому рішенні використовується пристрій IoT для аналізу звуків, які створюють машини, використовуючи розширені алгоритми штучного інтелекту та машинного навчання. Діагностика забезпечує оптимальне планування технічного обслуговування, скорочує дорогий час простою, підвищує операційну ефективність і прибутковість, а також оптимізує робочі процеси в різних галузях промисловості, включаючи агротехніку, виробництво, енергетику тощо. Крім того, компанія розробляє GEShI, або General Energy AI, який виконує функцію віртуального енергетичного радника. GEShI розроблено, щоб допомогти професіоналам у галузі енергетики орієнтуватися в складнощах енергетичного сектору за допомогою розширених можливостей, таких як прогнозна аналітика, прогноз споживання електроенергії та адаптивне навчання. Це дає професіоналам можливість впевнено приймати обґрунтовані рішення [20]. У березні 2024 року A-Gnostics AI укладає контракт із компанією Rodan Energy Solutions у Торонто, Канада. Прогнози енергоспоживання компанії успішно проникли на ринок Північної Америки, продемонструвавши вражаючу конкурентоспроможність порівняно з місцевими рішеннями з рівнем точності 98%.

Перевага використання нейронних мереж і машинного навчання полягає в тому, що комп'ютерні алгоритми можуть ідентифікувати різні фактори, які можуть бути неочевидними або суттєво змінилися, і обчислення виконуються значно швидше [19]. Людині може знадобитися значно більше часу, щоб знайти та зрозуміти такі зв'язки.

Ще одним важливим фактором є те, що Україна під час війни приєдналася до ENTSO-E [21], європейської мережі операторів систем передачі, яка об'єднує операторів 39 країн європейського континенту. Приєднання до інтегрованої мережі ще більше посилює переваги точного прогнозування. Пікові години споживання в Україні настають дещо раніше, ніж у Західній Європі, що дозволяє купувати електроенергію за нижчими цінами. І навпаки, коли попит в Україні падає, починається пік у Європі, що дозволяє здійснити продажі за вищими цінами. Штучний інтелект може швидше адаптуватися до цих умов.

## 2. ВИБІР МЕТОДІВ ПРОГНОЗУВАННЯ

### 2.1 Існуючі способи прогнозування електроенергії

Прогнозування використання електроенергії традиційно відбувається в два етапи:

- Створення математичної моделі для обробки даних про минуле використання.
- Використання моделі для складання прогнозу.

Від обраної математичної моделі залежить якість прогнозу. Є понад 100 модельних класів, але менше загальних класів.

Методи поділяються на дві групи: інтуїтивні та формалізовані. Інтуїтивні методи покладаються на експертну оцінку та професійний досвід для розуміння складних ситуацій, коли математичні моделі складні. Деякі методи включають експертні оцінки, історичні аналогії та прогнози на основі вибірки. [1] Поширеним є використання експертних систем і нечіткої логіки. Ці інтуїтивні методи досліджуються в статті.

Формалізовані методи включають моделі прогнозування.

#### **Метод авторегресії та ковзного середнього.**

У статистичному аналізі часових рядів моделі авторегресії та ковзного середнього (ARMA) забезпечують економічно ефективний опис (слабо) стаціонарного стохастичного процесу за допомогою двох поліномів: одного для авторегресії (AR/AP), а іншого для ковзного середнього (MA/KC). Загальна модель ARMA була вперше описана в дисертації Пітера Уіттла 1951 року «Перевірка гіпотез в аналізі часових рядів» і набула популярності завдяки книзі Джорджа Бокса та Гвіліма Дженкінса 1970 року [22].

Для даних часового ряду  $X_t$  модель ARMA служить інструментом для розуміння та потенційного прогнозування майбутніх значень ряду. Компонент AR передбачає регресію змінної на її власні відсталі (тобто минулі) значення [23]. Частина MA включає в себе моделювання члена помилки як лінійної комбінації членів помилки, що виникають у поточний момент і різні моменти в минулому. Цю модель зазвичай

називають моделлю ARMA(p, q), де p означає порядок компонента AR, а q представляє порядок компонента MA.

Відомо, що модель авторегресії має наступний вигляд [24]:

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \dots + \varphi_p z_{t-p} + a_t, \quad (1)$$

де  $z_t$  - центроване значення часового ряду у часовому моменті,  $t$ ,  $a_t$  - імпульс білого шуму з дисперсією  $\sigma_a^2$ , котрий був створений в той же момент  $t$ , можна, застосовуючи при цьому методику Юла-Уокера [22], згідно з якою числові значення коефіцієнтів регресії  $\varphi_i$ ,  $i = 1, 2, \dots, p$  знаходять із матричного рівняння

$$\varphi = M^{-1}\rho, \quad (2)$$

де матриця  $M^{-1}$  є оберненою до іншої матриці  $M$ , а наступні матриці  $\varphi$ ,  $M$ ,  $\rho$  виглядають наступним чином:

$$\varphi = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \dots \\ \varphi_p \end{bmatrix}, M = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \rho_2 & \dots & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \rho_1 & \dots & \rho_{p-3} \\ & & & \dots & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & & 1 \end{bmatrix}, \rho = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \dots \\ \rho_p \end{bmatrix} \quad (3)$$

В цій матриці  $\rho_i$ ,  $i = 1, 2, \dots, p$  - це автокореляції центрованого часового ряду  $z_t$ , які розраховують за виразом

$$\rho_i = \frac{\gamma_i}{\gamma_0}, i = 1, 2, \dots, p, \quad (4)$$

у якому  $\gamma_i$ ,  $i = 1, 2, \dots, p$  - автоковаріації, що вираховуються для центрованого часового ряду  $z_t$  за виразом:

$$\gamma_i = \frac{1}{N-i} \sum_{t=1}^{N-i} z_t z_{t+i}, i = 1, 2, \dots, p \quad (5)$$

та використовують, окрім розв'язання матричного рівняння (2), і для обчислення дисперсії  $\sigma_a^2$  білого шуму  $a_t$  наступним чином

$$\sigma_a^2 = \gamma_0 - \varphi_1 \gamma_1 - \varphi_2 \gamma_2 - \dots - \varphi_p \gamma_p, \quad (6)$$

у якій також можна підставити результати розв'язання матричного рівняння (2).

Вирази (1) - (6) створюють вихідні передумови для поставленої нами задачі побудови методу ідентифікації моделі авторегресії-ковзного середнього ARКС(p, q) з

задовільними значеннями порядків  $p, q$ , структура котрої для центрованого часового ряду  $z_t$  описується так

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \dots + \varphi_p z_{t-p} + a_t - \theta_q a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}, \quad (7)$$

шляхом подальшого розвитку методики Юла-Уокера.

До описаних вище 6 передумов необхідно додати ще одну – сьому, яка зумовлена відсутністю кореляції між сусідніми імпульсами білого шуму й може бути задана відомими виразами (1, 2):

$$\text{cov}(z_t a_{t-i}) = \begin{cases} \sigma_a^2, & \text{для } (i = 0) \\ 0, & \text{для } (i \neq 0) \end{cases} \quad (8)$$

$$\text{cov}(z_{t-k} a_{t-i}) = \begin{cases} \sigma_a^2, & \text{для } (i = k) \\ 0, & \text{для } (i \neq k) \end{cases} \quad (9)$$

На закінчення, щоб підкреслити нетривалість запропонованого методу ідентифікації моделі АРКС( $p, q$ ) із задовільними значеннями порядку  $p, q$  можна процитувати одну із сторінок 97 монографії Бокса і Дженкінса [24]: Як у випадку фіксованого  $\sigma_a^2$  тому і в іншому випадку фіксованого  $\sigma_\gamma^2$  (однакового за значенням  $\gamma_0$  – згідно з приміткою автора) оптимальний вибір призводить до деяких випадкових процесів, параметри яких є функціями невідомих динамічних параметрів. Таким чином, ми знаходимося у добре відомій парадоксальній ситуації, коли кращі дані можна зібрати, лише якщо щось уже відомо про шукану відповідь. Послідовний підхід, за якого ми вдосконалюємо режим у міру вивчення нових параметрів, відкриває можливості, які вимагають подальшого вивчення .

### **Експоненційне згладжування.**

Метод експоненціального згладжування — це послідовний метод, який згладжує параметри аналізованого ОУ за допомогою фільтра постійної пам'яті. Метод працює за таким принципом: попередні значення параметрів розглядаються з експоненціально зменшуваними вагами, дотримуючись певної моделі зважування (експоненціальним законом) [25].

$$s_t = \begin{cases} c_1: & t = 1 \\ s_{t-1} + \alpha \times (c_t - s_{t-1}): & t > 1 \end{cases} \quad (10)$$

де:  $s_t$  – згладжений ряд;  $c_t$  – первинний ряд;  $a$  – коефіцієнт згладжування, який обирається апріорі з діапазону ( $0 < a < 1$ ).

Для довільного скалярного параметра цей метод має такий вигляд:

$$\hat{\theta}_n = (1 - \xi)\theta_n + \xi f[\hat{\theta}_{n-1}, \hat{\theta}_{n-2}, \dots, \hat{\theta}_{n-s}], \quad (11)$$

де  $\theta_n$  – раніше виміряне значення параметра на  $n$ -кроці;  $\hat{\theta}_n$  – згладжене значення параметра в  $n$ -м огляді;  $f[\dots]$  – функція попередніх згладжених показників;  $\xi$  – стала величина, що має сенс коефіцієнта згладжування.

У найосновнішій формі метод використовує лише поточне виміряне значення та попереднє згладжене значення для визначення наступного значення параметра, яке потрібно оновити [26]. Отже, формулу (10) можна виразити так:

$$\hat{\theta}_n = (1 - \xi)\theta_n + \xi\hat{\theta}_{n-1}$$

Якщо ми візьмемо формулу під номером 2 та запишемо так:

$$\hat{\theta}_{n-1} = (1 - \xi)\theta_n + \xi(1 - \xi)\theta_{n-1} + \xi^2(1 - \xi)\theta_{n-2} + \dots + \xi^n(1 - \xi)\theta_0 = \sum_{i=0}^n \eta_i \theta_{n-i}, \quad (12)$$

де  $\eta_i = \xi^i (1 - \xi)$  – вагова функція, що враховує поточне та раніше виміряні значення параметрів об'єктів управління.

При згладжуванні параметра, який змінюється за лінійним законом, формула експоненційного згладжування приймає наступний вигляд:

$$\hat{\theta}_n = (1 - \xi)\theta_n + \xi f[\hat{\theta}_{n-1}, \hat{\theta}_{n-2}]. \quad (13)$$

У даному випадку операція  $f[\dots]$  над попередніми згладженими значеннями параметру має на меті визначити екстрапольоване значення параметра на момент  $t_n$  останнього вимірювання.

Суттєвою оцінкою якості згладжування може бути відношення середніх значень квадратів згладжених і незгладжених значень у разі, коли незгладжені дані являють собою дискретний випадковий процес з постійною спектральною щільністю [25]. Чим ближче це співвідношення до меншої одиниці, тим ефективніший процес згладжування; якщо це співвідношення більше одиниці, згладжування не досягнуто.

Отже, якщо незгладжені величини  $\theta$  описують собою дискретний випадковий процес з постійною спектральною щільністю  $S_{ff}^*(\bar{\omega}) = \sigma^2$  та з середнім значенням [27], що дорівнює нулю, то параметр квадрата цієї похибки дорівнює  $\bar{\theta}_n^2 = \sigma^2$

Це значення (середнє значення квадрата відхилення згладженої величини) можна знайти за такою формулою:

$$\bar{\theta}_n^2 = \frac{1}{\pi} \int_0^\pi |K_\mu^*(j\bar{\omega}, 0)|^2 S_{ff}^*(\bar{\omega}) d\bar{\omega}. \quad (14)$$

Розгорнутий вигляд формули (14)

$$\bar{\theta}_n^2 = \frac{1}{\pi} \int_0^\pi \left| \frac{1 - \xi}{1 + \xi[(1 - e^{-j\bar{\omega}})^{s+1} - 1]} \right|^2 \sigma^2 d\bar{\omega}. \quad (15)$$

Для того, щоб розрахувати цей параметр вдаємося до цієї формули:

$$P^*(q) = K_\mu^*(q, 0)\sigma = \frac{1 - \xi}{1 + \xi \sum_{v=1}^{s+1} (-1)^v \binom{s+1}{v} e^{-qv}} = \frac{(1 - \xi)e^{(s+1)q}}{e^{(s+1)q} + \xi \sum_{v=1}^{s+1} (-1)^v \binom{s+1}{v} e^{(s+1-v)q}}. \quad (16)$$

Використовуючи формулу для сумарної оцінки та вважаючи, що в ній  $l = s + 1$ , отримаємо:

$$\begin{aligned} b_k &= 1 - \xi \text{ при } k = s + 1 \\ b_k &= 0 \text{ при } k < s + 1 \end{aligned}$$

та  $a_{s+1} = 1$ ,  $a_{s+1-v} = (-1)^v \binom{s+1}{v}$ , де  $(v = 1, 2, \dots, s+1)$ . Після цього отримуємо після обчислення для різних  $s$ :

$$\left\{ \begin{aligned} s = 0: & \frac{\bar{\theta}_n^2}{\sigma^2} = \frac{1 - \xi}{1 + \xi}; \\ s = 1: & \frac{\bar{\theta}_n^2}{\sigma^2} = \frac{1 + \xi}{1 + 3\xi}; \\ s = 2: & \frac{\bar{\theta}_n^2}{\sigma^2} = \frac{(1 - \xi)(1 + 4\xi)}{(1 + \xi)(1 + 7\xi)}; \\ s = 3: & \frac{\bar{\theta}_n^2}{\sigma^2} = \frac{1 + 8\xi - 25\xi^2}{(1 - 5\xi)(1 + 15\xi)}; \end{aligned} \right. \quad (17)$$

В цих формулах  $\xi$  не повинно бути більшим за граничне значення [26].

Основними перевагами цього методу є його простота та легкість застосування, а також ряд інших переваг. Ось найактуальніші з них:



- Він вимагає мінімальних історичних даних порівняно з іншими методами, такими як ARIMA.
- Він забезпечує вищу точність, ніж інші методи, коли використовується експоненціальне моделювання.
- Це дуже гнучкий метод, який дозволяє дослідникам вибирати дані попиту, з якими вони працюють.
- Техніка подвійного експоненціального згладжування допомагає пом'якшити проблеми прогнозування, коли коефіцієнт згладжування перевищує 0,5, усуваючи один із небагатьох недоліків [27].

### **Метод найближчого сусіда.**

Важливо зазначити, що метод «найближчого сусіда» відноситься до категорії алгоритмів, які покладаються на збереження даних у пам'яті для порівняння з новими елементами. Обчислювальна складність алгоритму —  $O(n^2)$  [28]. Коли з'являється новий запис для прогнозування, алгоритм обчислює відхилення між цим записом і подібними наборами даних, згодом ідентифікуючи найбільш схожий запис або найближчого сусіда. У цьому підході використовується термін « $k$ -найближчі сусіди», який означає, що «верхні» (найближчі) сусіди вибираються як набір «найближчих сусідів» для розгляду. Оскільки зберігати всі дані не завжди практично чи зручно, іноді зберігається лише колекція «типових» випадків. У таких випадках метод, який використовується, називається Case Based Reasoning (CBR) або міркуванням за аналогією, яке передбачає міркування на основі подібних випадків або прецедентів [29].

Алгоритм найближчого сусіда можна описати так:

1. Треба почати з будь-якої довільної точки в наборі точок  $V$ , де  $V$  є набором  $N$ -вимірних точок.
2. Далі переходити до найближчої точки, яка раніше ще не була досліджена. Включити відвідані точки в маршрут одну за одною, та переконатися, що наступна точка, додана до маршруту, є найближчою до останньої вибраної точки серед усіх невідвіданих точок [29].
3. Продовжувати цей процес, доки всі точки в наборі  $V$  не будуть відвідані.

4. З'єднати останню точку маршруту з першою, утворюючи повний маршрут.

5. Вхідні дані: набір точок  $V$  розмірністю  $N$  вимірами. Вихідні дані: побудований маршрут  $T$ , що складає собою послідовність відвідування точок множини  $V$ .

Формула розрахунку прогностичних значень за цим методом:

$$m_n = \sum_{x(i)} w(x - X_{(j)}X_{(j+1)}), \quad (18)$$

де  $w$  – залежність, що прив'язується до сусідів параметром  $k$

Спосіб задання параметра  $w(x - X_{(j)})$  через коефіцієнт  $k$  описується такою формулою:

$$w(x - X_{(j)}) = \frac{\frac{1}{n}}{\frac{1}{n} \sum_{i=1}^n k \frac{x - X_{(j)}}{R_n}}, \quad (19)$$

де  $R_n$  – відстань між  $x$  та найближчими сусідами.

Шлях обрахунку коефіцієнта  $k$  визначається таким способом [30]:

$$k = \exp(-\|x - X_{(j+1)}\|)^2. \quad (20)$$

**Переваги методу:**

- Простота використання, де отримані результати легко застосовувати.
- Універсальність: рішення не стосуються окремої ситуації й можуть використовуватися в інших випадках.

- Пошук має на меті знайти найкраще можливе рішення, а не гарантовано правильне.

**Недоліки методу «найближчого сусіда»:**

- Метод не створює моделей або правил, які узагальнюють попередній досвід і приблизно на 25% маршрут буде побудований довше від оптимальнішого. Натомість він покладається на весь набір історичних даних, що унеможлиблює розуміння основи для відповідей [23].

- Труднощі у виборі міри «близькості» (метрики): вибір метрики в основному залежить від розміру набору записів, який буде зберігатися в пам'яті для задовільної класифікації або передбачення. Результати класифікації сильно залежать від вибраного показника.

- Він вимагає повного пошуку навчальної вибірки під час розпізнавання, що призводить до збільшення обчислювальної складності.

- Метод зазвичай стикається з проблемами з малими розмірами з точки зору кількості класів і змінних, що призводить до обмеженого застосування.

Нижче у таблиці 1 наведені приклади вищезгаданих та інших моделей прогнозування. Оціночними критеріями тут виступають прозорість, адаптивність, гнучкість, врахування треду та здатність моделювати нелінійні процеси. Кожен метод має свої мінуси та плюси, але із 8 кандидатів виділяються тільки 2: моделі групового врахування аргументів та Хольта-Вінтерса.

Таблиця 1 – Порівняння методів прогнозування

Моделі	Гнучкість	Прозорість	Адаптивність	Здатність моделювати нелінійні процеси	Врахування треду
Авторегресійна ковзного середнього	+	+	-	-	+ -
Експоненціального згладжування	-	+	+	+	-
Хольта	-	+	+	+	+ -
Хольта-Вінтерса	-	+	+	+	+ +
Групового врахування аргументів	+	-	+	+	+ +
Лінійна регресійна	+	+	-	-	+ -
Нелінійна регресійна	+	+	-	-	+ -

Множинна регресійна	+	+	-	-	+ -
------------------------	---	---	---	---	--------

## 2.2 Нечіткий регресійний аналіз

Методи регресійного аналізу часто використовуються для побудови моделі з використанням вибірки зібраних статистичних даних і отримання прогнозного рівняння для повного набору даних. Побудова регресійних моделей може бути легко полегшена за допомогою різних пакетів прикладного програмного забезпечення на комп'ютері. Але обчислювальні машини можуть обробляти лише звичайні точні числа, що ускладнює використання якісних або вербальних даних для побудови регресійних моделей. У реальних проблемах спрощення даних може призвести до ігнорування важливої інформації для регресійних моделей [7]. Іноді таке спрощення є непрактичним і може вплинути на застосовність отриманих ресурсів. Певні спостереження, наприклад дані, надані експертами, можна описати лише в лінгвістичній формі (наприклад, задовільно, добре або відмінно). Для них теорія нечітких множин пропонує апарат для обробки лінгвістичних значень за допомогою нечітких функцій належності.

Для врахування нечітких даних при побудові регресійних моделей запропоновано НРА. На відміну від звичайного регресійного аналізу, який ґрунтується на теорії ймовірностей, нечіткий регресійний аналіз може базуватися на теорії можливостей і теорії нечітких множин [31]. У традиційному регресійному аналізі нескориговані помилки між регресійною моделлю та спостережуваними даними зазвичай вважаються порушеннями спостереження, які є випадковими величинами з нормальним розподілом і математичним сподіванням, рівним нулю. У нечіткому регресійному аналізі ці самі нескориговані відхилення пояснюються невизначеністю структури моделі, як запропоновано Такака [8]. На сьогоднішній день розроблено різні методи нечіткої регресії з використанням кількох критеріїв оптимальності для визначення нечіткого центру та розкидів .

Враховуючи вищезазначену інформацію, для відповідності моделі нечіткої регресії використовуються два основні підходи:

**Можлива модель** спрямована на мінімізацію нечіткості моделі шляхом мінімізації загальних розкидів її неявних коефіцієнтів. Це підлягає обмеженню, що точки даних кожного зразка повинні входити в заданий можливий інтервал даних.

**Модель найменших квадратів:** цей підхід спрямований на мінімізацію відстані між результатом моделі та спостережуваним результатом. Він враховує їх моделі та розповсюдження.

Лінійна регресійна модель досліджуваної системи, запропонована Шапіро, представлена лінійною комбінацією значень її вхідних змінних

$$Y^*(x_j) = A_0 + A_1x_{1j} + \dots + A_{nj}x_{nj}. \quad (21)$$

Звичайна регресійна модель передбачає, що системна характеристика визначається різкими і точними значеннями, а будь-які відхилення між спостережуваними і розрахунковими варіантами залежних змінних обумовлені похибками спостереження [8]. Однак статистичні регресійні моделі, засновані на принципах теорії ймовірностей, є точними лише за умови дотримання певних передумов, як зазначають Покорний (1993) та Шапіро (2006). Деякі загальні практичні питання включають:

- Невелику кількість спостережень або розмірів вибірки, який занадто обмежений.
- Неможливість гарантувати нормальний розподіл помилок.
- Складність у визначенні зв'язку (невизначеності) між вхідними і вихідними змінними.

Цих проблем можна уникнути при створенні регресій з використанням теорії можливостей і виявлення регресійної залежності як нечіткої функції. Відхилення між спостережуваними та оціночними варіантами залежних змінних не може бути значно викликано поганими локальними змінними структури системи [7]. Замість цього ці варіації можуть бути обумовлені не дуже гострим характером параметрів системи.

Такі нечіткі явища також повинні відображатися в нечіткості відповідних параметрів моделі.

При розгляді нечіткості регресійної моделі можна розглянути два випадки (які не є взаємовиключними). По-перше, ми припускаємо, що вхідні дані є точними, а невизначеність полягає у визначенні моделі. У цьому випадку неясність відображається нечітким характером коефіцієнтів регресії як параметрів моделі. У другому випадку ми можемо вважати систему чітко визначеною, але виміряні дані мають неточний характер. Отже, невизначеність моделі здійснюється невизначеними вхідними даними [31].

### **Модель інтервальної лінійної регресії**

Перший значний крок у створенні нечіткої регресійної моделі зробив Баклі [33], який розробив технологію інтервальних регресійних моделей (вищезгадана формула Шапіро (21), але інші позначення)

$$\check{Y}(x_j) = \check{A}_0 + \check{A}_1 x_{1j} + \dots + \check{A}_j x_{nj}. \quad (22)$$

У цьому підході  $\check{Y}(x_j)$  представляє оціночне значення вихідної змінної, як замкнутий числовий інтервал, що відображає невизначеність неспецифічної системи. Коефіцієнти регресії моделі, позначені  $\check{A}$ , також представлені у вигляді розпливчастих замкнутих числових інтервалів. Для визначення інтервалів коефіцієнтів регресії використовується метод лінійного програмування, описаний в Каспрзык (1992) [34]. Для алгебраїчних обчислень з інтервальними числами проста арифметика інтервалу була розроблена Муром (1979) [35].

### **Модель нечіткої лінійної регресії**

Наступним кроком у розвитку невизначених регресійних моделей є створення моделей, які використовують неясність, формалізуючи невизначеність неточними інтервалами замість числових інтервалів. Ці регресійні моделі, які фіксують невизначеність модельованих систем, називаються нечіткими регресійними моделями (Каспрзык, Полещук, Шапіро). Невизначений характер моделі нечіткої регресії представлений розрахунковими значеннями неточного виходу, що позначаються як  $\check{Y}^*(x_j)$ , і коефіцієнтами нечіткої регресії, представленими як  $\check{A}$ , у вигляді спеціалізованих

неявних множин, що називаються неоднозначними числами. Форма моделі нечіткої лінійної регресії описана в роботах Баклі (2008), Хешматі (1985) і Танаки (1982) [36].

Невизначений характер моделі нечіткої регресії представлений розрахунковими значеннями неточного виходу  $\tilde{Y}^*(x_j)$  і коефіцієнтами нечіткої регресії  $\tilde{A}$  у вигляді спеціалізованих розмитих множин, відомих як нечіткі числа. Форма моделі нечіткої лінійної регресії, як описано Баклі, Хешматі і Танака, задається рівнянням:

$$\tilde{Y}^*(x_j) = \tilde{A}_0 + \tilde{A}_1 x_{1j} + \dots + \tilde{A}_j x_{nj} = \tilde{A} \cdot x', \quad (23)$$

де  $x'$  - це транспонований вектор стовпця  $x = (x_1, x_2, \dots, x_n)$ , і вектор параметрів, елементами якого є нечіткі числа [32]. У функції нечіткої регресії,  $\tilde{A}$  є багатовимірною невизначеною множиною (нечітким співвідношенням) в результаті декартового добутку нечітких множин та розмитих параметрів

$$\tilde{A} = \tilde{A}_0 \times \tilde{A}_1 \times \dots \times \tilde{A}_n \quad (24)$$

з функцією належності у формі

$$\mu_{\tilde{A}}(a) = \bigcup_{i=1}^n \{\mu_{\tilde{A}_i}(a)\}, \quad a = (a_1, a_2, \dots, a_n). \quad (25)$$

Форма функції залежностей для вихідних значень нечітких чисел в моделі нечіткої лінійної регресії (21) розраховується Zadeh's на основі принципу розширення Novak [37] в наступному вигляді:

$$\mu_{\tilde{Y}}(y) = \bigcup_{a|at'=y} \mu_{\tilde{A}}(a); \{a|t' = y\} = 0. \quad (26)$$

Функція приналежності  $\mu_{\tilde{A}_i}(a_i)$  апроксимується у вигляді трикутних нечітких чисел, як описано в Новаку (1990) [37]:

$$\mu_{\tilde{A}_i}(a_i) = 1 - \frac{|a_i - a_i|}{c_i}; \quad a_i - c_i \leq a_i \leq a_i + c_i, \quad (27)$$

де  $a_i$  - середнє значення (ядро) нечіткого числа  $\tilde{A}_i$ , а  $c_i$  - половина ширини носія, що підтримує  $\tilde{A}_i = \{a_i, c_i\}$ . Функції приналежності для вихідних нечітких множин (23) можуть бути виражені в наступному вигляді, як описано Каспрзык [34]:

$$\mu_{\tilde{Y}}(y) = 1 - \frac{|y - a \cdot x'|}{\sum_{i=1}^n c_i |x_i|}; \quad a \cdot x' - \sum_{i=1}^n c_i |x_i| \leq y \leq a \cdot x' + \sum_{i=1}^n c_i |x_i|, \quad (28)$$

## 2.3 Інформаційні технології прогнозування електроенергії

Інформаційні технології значно змінили сферу прогнозування ЕЕ. Зі зростанням попиту на надійну та стійку енергію, оптимізацію витрат та енергоефективність, інтеграція передових ІТ-рішень стала вирішальною для вирішення складних проблем сучасного сектору [1].

Одним з основних способів впливу даного розділу є використання великих даних та аналітики в цій сфері, надаючи потужні інструменти і методи для більш точних і надійних прогнозів споживання енергії. Вживання ІТ-систем дозволило збирати та обробляти величезну кількість відомостей з різних джерел, які можуть бути проаналізовані для виявлення тенденцій та закономірностей, які можуть допомогти в передбаченні майбутнього попиту на ЕЕ [19].

Із первинних джерел даних, для подальшого проектування, є механізм розумної енергосистеми. Вона представляє собою модернізацію енергетичної інфраструктури, котра експлуатує датчики, автоматизацію та інші цифрові технології для підвищення ефективності та надійності. Інформаційно-технологічні моделі можуть збирати факти про споживання в режимі реального часу з інтелектуальних лічильників та історичні - з традиційних [38]. Ці дані можуть бути застосовані для аналізу механізмів попиту та виявлення факторів, які впливають на запит електроенергії, таких як час доби, дня тижня та сезонні зміни, контролювати потреби та постачання online, що запобігає відключенню струму та скороченню викидів парникових газів.

Погодні відомості також представляються важливим матеріалом інформації. Системи збирають дані про температуру, вологість, вітер і сонячну активність, котрі пізніше задіюються для прогнозування змін зацікавленості на ЕЕ через погодні умови. Наприклад, у спекотні літні місяці попит на електроенергію має тенденцію до збільшення через використання кондиціонера, тоді як у холодні зимові місяці попит може збільшуватися через використання систем опалення, але різниця експлуатації в даних прикладах може бути різною [39]. Поновлювані джерела енергії, такі як гідро та геотермальна енергія, стають все більш популярними через менший вуглецевий слід та стійкість у порівнянні з традиційним викопним паливом. Ці енергетичні



потоки також перетворюються у більш доступні, роблячи їх досяжними як для великого, так і для малого бізнесу.

Системи зберігання енергії є ще однією важливою технологією в енергетичному секторі. Вони використовуються для відкладу ВДЕ і традиційних джерел з метою забезпечення стабільного постачання струму [39]. Акумуляторні моделі зберігання в даний час є найбільш поширеним рішенням, що дозволяє консервувати енергію в години поза піком і вивільняти під час пікового запиту.

Розробка Internet of Things (IoT) стала на допомогу вищезгаданому сектору для оптимізації та зменшенню відходів. Пристрої IoT, такі як інтелектуальні термостати та енергоефективне освітлення, можуть бути підключені до Інтернету та контролюватися дистанційно. Вони запрограмовані на online-режим, щоб контролювати використання енергії та підвищувати ефективність [38].

Штучний інтелект - ще один інструментарій, який дозволяє еволюціонувати енергетичному сектору [17]. Алгоритми AI можуть аналізувати дані для прогнозування потреби, оптимізації використання та підвищення ефективності. Це також можна використовувати для поновлюваних джерел, пророкуючи виробництво енергії вітру та сонця.

Інші критерії фокусуються на зниженні вартості або складності, пов'язаної з виробництвом, продажем, дистрибуцією, підтримкою та іншими видами діяльності в рамках ланцюжка цінової мітки. По суті, інформаційні технології передбачування - це постійний пошук шляхів підвищення цінності продукту або послуги як для своїх клієнтів, так і для кінцевих користувачів, а також для суб'єктів, що беруть участь у його розробці, виробництві, маркетингу та підтримці [40]. Під час цього процесу визначаються системи, підсистеми та компоненти, які будуть інтегровані в товар, завжди пам'ятаючи про важливість максимального значення для всіх сторін, що беруть участь.

Об'єктами аналізу, спрямованого на прогнозуванні методики, є технологія, продукт, компанія, промисловість та країна. Кожен з цих сутностей відіграє вирішальну роль у формуванні ринку в цілому, і їх аналіз може допомогти в прийнятті обґрунтованих рішень для розвитку і використання цих технік.

Першим об'єктом аналізу в прогнозуванні технологій є сама **технологія**. Більшість досліджень спираються на єдину патентну базу даних для технічного аналізу або використовують базу даних з патентами та публікаціями з різних джерел для виявлення процесуальних можливостей. Технічна експертиза передбачає вивчення особливостей і можливостей технології, її потенційних застосувань та її обмежень [40]. Ця інформація може бути використана для прогнозування майбутніх розробок і тенденцій.

**Продукт.** Деякі вивчення відокремлені від єдиного технічного аналізу, а інноваційний метод продукту вибирається залежно від економічної та технічної сили компанії. Аналіз товару передбачає розвідку особливостей та можливостей виробництва, його потенційного ринку та конкуренції. Ця інформація може бути використана для прогнозування майбутніх розробок і тенденцій [6].

**Компанія.** Частина досліджень зосереджена на зборі та аналізі точного, перспективного та дієвого інтелекту про бізнес-середовище, суперників та саму організацію на рівні корпорації. Це може допомогти менеджерам оцінити конкурентів і постачальників для підвищення конкурентоспроможності та ефективності. Аналіз підприємства передбачає вивчення сильних, слабких сторін, можливостей та загроз (SWOT) [4] фірм, її фінансових показників та стратегічного напрямку.

Четвертий розділ займає - **галузь**. Галузевий аналіз передбачає вивчення загальних тенденцій ринку, конкуренції та регуляторного середовища сектору. Інформація може бути використана для прогнозування майбутніх трендів розвитку блоку [1]. Джерела даних для аналізу промисловості включають патенти, що містять технічні знання, товарні знаки з відомістю про продукти та онлайн-форуми з повідомленнями про ринок.

П'ятий об'єкт аналізу - **країна**. Ревізія передбачає вивчення економічних, політичних і соціальних умов країни, які можуть вплинути на технологічний розвиток і використання. Ця інформація може бути описана для прогнозування майбутніх технологічних розробок і тенденцій в конкретній країні. Джерела даних для розгляду країни включають патенти, публікації та економічні та демографічні дані [8].

На закінчення, інформаційна технологія передбачення є складною системою, яка включає в себе ревізію різних об'єктів для прогнозування майбутніх технічних розробок і тенденцій. Об'єктами аналізу, спрямовані на ці промисли, є технологія, продукт, компанія, промисловість та країна [40]. Кожен з них відіграє вирішальну роль у формуванні суспільства в цілому, і їх оцінка може допомогти в прийнятті обґрунтованих рішень про розвиток і використання інновацій. Вивчаючи логічний зв'язок між цими аспектами, антиципація технологій може забезпечити цінне розуміння майбутніх технологічних виготовлень та трендів.

### 3. ВИБІР ОПТИМАЛЬНОГО МЕТОДУ ПРОГНОЗУВАННЯ

#### 3.1 Порівняння та аналіз моделей прогнозування

Для ефективного управління енергоспоживанням і витратами сьогодні необхідна надійна модель прогнозування. Представлені моделі підкреслюють три основні аспекти: споживання, генерація та дані про витрати енергії. Мета полягає в тому, щоб спрогнозувати майбутнє використання ЕЕ, виробничих потужностей і відповідних витрат шляхом ретельного вивчення історичних даних і асиміляції поточних тенденцій. Цей всеохоплюючий метод забезпечує більш точне та повне розуміння енергетичного середовища підприємства, покращуючи можливості прийняття рішень та стратегічного планування. Враховуються також різноманітні фактори, такі як сезонні коливання, робочі зміни, коливання ринку та зміни політики, щоб забезпечити максимально прагматичний і надійний прогноз.

У цьому розділі розглядається аналіз різних методів, які застосовуються до прогнозування споживання електроенергії на основі ІТ. Увага буде зосереджена на підходах, котрі можуть ефективно впоратися з невизначеністю для створення надійних і адаптованих прогнозів. Переглядаючи сильні та слабкі сторони різних моделей, необхідно буде визначити найбільш прийнятні рішення для підприємств, які прагнуть отримати кращий контроль над споживанням енергії та витратами.

Статистичні моделі, методи машинного навчання та економетричні методи є основними категоріями, які охоплюють широкий спектр конкретних практик прогнозування, придатних для споживання електроенергії в умовах невизначеності [1]. Вони є фундаментальними стовпами для вирішення даної проблеми. Також можна використовувати гібридні моделі, наприклад, статистичні методи для базового прогнозування та машинне навчання для фіксації нелінійних зв'язків із даними про погоду.

#### Статистичні моделі

Компаніям необхідно оптимізувати витрати на енергію, планувати майбутні потреби та укладати кращі контракти. Тому дана проблема має вирішальне значення

для існування бізнесу та подальшу роботу у цьому секторі. Однак невизначеності, такі як коливання графіків виробництва, погоди та ринкова динаміка, можуть значно вплинути на точність результату. СМ пропонують потужний набір інструментів для вирішення цієї проблеми.

Статистичні моделі — це математичні системи, які використовують історичні дані та статистичні методи для прогнозування. Вони покладаються на припущення, що в даних існують закономірності, що дозволяє оцінити майбутню поведінку на основі минулих спостережень. У контексті прогнозування споживання електроенергії ці моделі аналізують історичні дані про використання поряд із потенційно значущими факторами, такими як погода, рівень виробництва та графіки зайнятості [41].

Сюди входять такі види моделей:

- **Авторегресійне інтегроване ковзне середнє (ARIMA) та експоненціальне згладжування.**
- **Регресійний аналіз** - встановлює зв'язок між залежною змінною (наприклад, споживанням електроенергії) та однією або кількома незалежними змінними (наприклад, температурою, обсягом виробництва). Це допомагає визначити, як ці фактори впливають на споживання, і включає їх у прогноз.
- **Сезонна авторегресійна інтегрована ковзаюча середня (SARIMA)** - є розширенням моделі ARIMA, яка спеціально враховує сезонні закономірності в даних (погода, святкові періоди).
- **Фільтри Калмана** - це рекурсивні методи оцінювання, які поєднують шумові вимірювання датчика з динамічною моделлю для отримання більш точної оцінки стану. У прогнозуванні споживання електроенергії їх можна використовувати для відстеження моделей використання в реальному часі та включення їх у прогноз, навіть якщо дані датчиків є невизначеними.
- **Моделі GARCH (узагальнена авторегресійна умовна гетероскедастичність)** - це клас моделей, що використовуються для моделювання змінної дисперсії у часових рядах фінансових даних. Вони особливо корисні для прогнозування волатильності на фінансових ринках. Моделі GARCH дозволяють узагальнити

класичні моделі ARMA, додаючи до них здатність моделювати змінну дисперсію, що змінюється в часі.

Вони надають можливість інтерпретації: пропонують чітке уявлення про те, які історичні фактори та закономірності впливають на прогноз. Це дозволяє легше діагностувати потенційні проблеми та вдосконалювати техніку робочого процесу [41]. Також не потребують великої обчислювальної потужності порівняно зі складними методами машинного навчання, що робить їх придатними для ситуацій з обмеженими ресурсами. І один із самих головних плюсів – це ретельне вивчення та застосовування протягом десятиліть. Велика купа досліджень, робіт, статей надає доступні способи та добре зрозумілі обмеження.

Хоча СМ не можуть усунути невизначеність, вони пропонують кілька способів її включення та обліку. Наприклад, довірчі інтервали, які кількісно визначають діапазон, у межах якого фактичне значення ймовірно потраплятиме з певним рівнем довіри. Це якраз і допомагає повідомити про притаманну невизначеність, пов'язану з прогнозом. Також деякі СМ, наприклад ARIMA з екзогенними змінними (ARIMAX), можуть враховувати вплив зовнішніх факторів, таких як погода, з імовірнісним компонентом. Вони називаються стохастичними моделями та включають в прогноз невизначеність.

Але ці методи також мають свої недоліки. Перший із них – це залежність від даних. Бо саме точність результату значною мірою залежить від якості та повноти історичних відомостей. Неточності або відсутність деталей можуть надати ненадійні прогнози [22]. Це призводить до труднощів з адаптацією: їм важко адаптуватися до раптових змін у моделях споживання, викликаних непередбаченими подіями або значними змінами в бізнес-операціях. Ще однією перешкодою є те, що СМ часто мають проблеми з дуже нелінійними зв'язками між змінними, які можуть переважати в даних про споживання електроенергії під впливом різноманітних факторів.

Статистичні моделі є цінною основою для підприємства в умовах невизначеності. Їх можливість інтерпретації, ефективність і усталені методи роблять їх добре підходящими для різних сценаріїв. Використовуючи методи кількісної оцінки невизначеності та вивчаючи різні статистичні підходи, підприємства можуть

використовувати ці моделі для отримання цінної інформації для прийняття обґрунтованих рішень щодо управління енергією.

### **Машинне навчання**

Машинне навчання охоплює широкий спектр алгоритмів, які навчаються на основі даних без явного програмування. Ці алгоритми визначають закономірності та зв'язки в даних, що дозволяє робити прогнози щодо нових, невідомих точок даних. У контексті прогнозування споживання електроенергії моделі машинного навчання можуть аналізувати історичні дані про використання ресурсів поряд з різними відповідними факторами, як-от погода, рівень виробництва та графіки зайнятості. Навчаючись на цих закономірностях, модель може передбачити майбутнє споживання навіть із внутрішньою невизначеністю [17].

Сюди відносяться такі моделі:

- **Штучні нейронні мережі (ШНМ)** засновані на моделюванні структур, ШНМ складаються з взаємопов'язаних вузлів (нейронів), які обробляють інформацію шарами. Ці мережі чудово вловлюють складні нелінійні зв'язки між змінними, що робить їх добре придатними для прогнозування проблем із різноманітними факторами впливу.
- **Опорновекторні машини (SVM)** спрямовані на пошук оптимальної гіперплощини у просторі великої розмірності, яка найкраще розділяє точки даних, що належать до різних класів. У прогнозуванні це можна використовувати для класифікації періодів високого та низького споживання на основі історичних даних і факторів впливу.
- **Дерева рішень** використовується для розв'язання задач класифікації та регресії. Він базується на створенні дерева з рішеннями, де кожен вузол представляє питання про характеристику даних, а кожне гілля відповідь на це питання. Дерева рішень легко інтерпретувати та використовувати, а також добре пристосовані до роботи з якісними та кількісними даними.
- **Випадкові ліси** - це метод ансамблю, котрий поєднує кілька дерев рішень, де кожне дерево робить прогноз на основі підмножини характеристик. Остаточним

прогнозом є середнє або бїльшїсть голосїв окремих дерев. Цей пїдхїд допомагає пїдвищити точнїсть і надїйнїсть порївняно з одним деревом рїшень.

- **Бассовї мережї** - це графїчний пїдхїд моделювання для вїдносин ймовїрностей мїж змїнними. Вони використовують теорему Басса для оновлення ймовїрностей на основї нових даних або спостережень. Цї мережї застосовуються в штучному їнтелектї, медицинї, фїнансах та їнших галузях для моделювання вїд невизначеностї та прийняття рїшень.
- **Гауссовї процеси** - сїмейство стохастичних процесїв, в яких будь-яка скїнченна кїлькїсть змїнних мають гауссївський розподїл. Вони широко застосовуються в багатьох галузях, включаючи фїнанси, машинне навчання, обробку сигналїв та природних мов, для симуляцїї неперервних випадкових процесїв.
- **Генетичнї алгоритми** - це метаевристика пошуку, що їмітує процеси природного вїдбору та еволюцїї. Вони використовуються для розв'язання складних оптимїзацїйних задач, шукання найкращих рїшень у просторї можливих рїшень.
- **Функцїї переконань** - є математичним їнструментом для моделювання та оцїнки невизначеностї в прийняттї рїшень. Функцїї переконань дозволяють враховувати ступїнь впевненостї у рїзних сценарїях, що допомагає у прийняттї кращих рїшень в умовах невизначеностї.

Вони можуть ефективно фїксувати складнї нелїнїйнї зв'язки мїж змїнними, якї часто присутнї в даних про споживання електроенергїї пїд впливом рїзноманїтних факторїв (погода, графїк виробництва). Це забезпечує значну перевагу перед традицїйними статистичними моделями. Алгоритми постїйно навчаються та адаптуються до нових моделей даних (це має вїршальне значення для ситуацїй, коли моделї споживання або дїлові операцїї змїнюються). А це в свою чергу створює новї функцїї на основї наявних вїдомостей, якї можуть бути бїльш їнформативними для моделї. В кїнцї МН отримує глїбшу їнформацїю зї складних наборїв даних.

Як ї статистичнї моделї, машинне навчання також не може усунути невизначенїсть, але має свої способи включити та врахувати їх: методи Байєса розглядають параметри моделї як розподїли ймовїрностей, що дозволяє включати



попередні знання про невизначеність у її прогнози [17]. Ансамбльове прогнозування поєднує передбачення із кількох методів машинного навчання з різними архітектурами (наприклад, ШНМ, випадкові ліси), яке призводить до більш надійного кінцевого результату і враховує потенційні невизначеності в кожній окремій моделі. А регуляризація вилучення відкидає нейрони під час навчання, запобігає надмірному підбору шаблону до навчальних даних і покращує її здатність узагальнювати невидимі дані з притаманною невизначеністю.

Мінуси МН відрізняються від вищеперерахованих моделей і мають багатенько своїх недоліків. Вимоги до даних, які ставляться перед цими моделями, часто потребують обсягів великої якості інформації для ефективного навчання. Це може становити проблему для підприємств з обмеженими історичними даними. При порівнянні зі статистичними моделями, особливо складними (такими як глибокі нейронні мережі) їхнє навчання може мати меншу можливість інтерпретації. Це погіршить розуміння того, як модель формує свої прогнози та ускладнить зусилля з виявлення можливих упереджень. А сама вартість обчислень для тренування може бути великою з точки зору обчислювальних витрат, і потребувати високопродуктивних обчислювальних ресурсів [42].

Після аналізу моделей машинного навчання, можна зрозуміти, що вони пропонують потужний набір інструментів для заданої теми. Сильна сторона його методів полягає у фіксуванні нелінійних зв'язків, адаптації до змінних шаблонів даних і врахуванні невизначеності за допомогою різних методів. Однак необхідно враховувати такі обмеження, як вимоги до даних, можливість інтерпретації та витрати на обчислення. Ретельно відбираючи та розгортаючи моделі машинного навчання, підприємства можуть отримати цінну інформацію для оптимізації споживання енергії.

### **Економетричні методи**

Експертні системи (інформують прогнози, але не прогнозують безпосередньо) - можуть відігравати допоміжну роль, хоча і не підпадають під основні категорії прогнозування споживання електроенергії [41]. Досвід експертів, їхні роботи та знання допомагають удосконалити існуючі моделі, розробити правила і порогові

значення, пояснити і обґрунтувати міркування. Вдосконалені ЕМ можуть бути інтегровані з способами прогнозування для створення більш комплексної системи підтримки прийняття рішень. Загалом, хоча ЕМ не є автономними методиками, вони діють як цінні сховища знань і можуть підвищити загальну ефективність підходів до передбачення споживання електроенергії підприємством.

На рисунках 3.1 та 3.2 наглядно показана таблиця де порівнюються різні відомі методи із кожної вищеперерахованої категорії. Там включаються як сильні так і слабкі сторони, інтерпретованість, вимоги до даних та інше. В ній можна провести аналіз та обрати саме підходящу форму для даної задачі.

1	Метод	Категорія	Сильні сторони	Слабкі сторони	Вимоги до даних	Інтерпретованість	Вартість розробки	Придатність до невизначеності
2	ARIMA	Статистичний	Уявлення про фактори, що впливають на прогнози.	Обмежена складність: можуть виникати проблеми з дуже нелінійними даними.	Загалом нижче	Висока	Відносно мала	Може бути придатним із відповідними техніками.
3			Потрібна менша обчислювальна потужність.	Залежність даних: точність залежить від якості даних.				
4			Добре зрозумілі методи з легкодоступними інструментами.	Труднощі з адаптацією до раптових змін.				
5	Експоненціальне згладжування	Статистичний	Потрібна менша обчислювальна потужність.	Обмежена здатність фіксувати складні відносини.	Загалом нижче	Висока	Відносно мала	Може бути придатним із відповідними техніками.
6			Добре підходить для короткострокового прогнозування.	Не ідеально підходить для довгострокового прогнозування або фіксації сезонності.				
7	Регресійний аналіз	Статистичний	Визначає зв'язки з факторами впливу.	Обмежена здатність працювати з нелійними зв'язками.	Помірна	Помірна	Помірна	Може бути придатним із відповідними техніками.
8			Можна використовувати для довгострокового прогнозування з додатковими міркуваннями.					
9	Штучні нейронні мережі (ШНМ)	Машинне навчання	Вловлює складні нелінійні зв'язки між змінними.	Вимоги до даних: часто потрібні великі обсяги високоякісних даних.	Загалом вище	Може бути низьким для складних моделей.	Більше часу та досвіду	Потужний із відповідними техніками.
10			Адаптується до шаблонів даних, що розвиваються.	Можливість інтерпретації: складні моделі можуть бути складними для розуміння.				
11	Машини підтримки векторів (SVM)	Машинне навчання	Ефективний для даних великої розмірності.	Може бути обчислювально дорогим для великих наборів даних.	Помірна	Помірна	Помірна	Може бути придатним із відповідними техніками.
12			Визначає ключові характеристики для прогнозування.	Може знадобитися досвід налаштування параметрів.				

Рисунок 3.1 – Порівняльна таблиця методів прогнозування (1)

14	Випадкові ліси	Машинне навчання	Міцний до викидів і переобладнання.	Може бути обчислювально дорогим для великих наборів даних.	Помірна	Помірна	Помірна	Може бути придатним із відповідними техніками.
15			Надає розуміння важливості функцій.	Здатність інтерпретації може бути нижчою порівняно з більш простими моделями.				
16	Модель нечіткої логіки	Статистичний	Може впоратися з невизначеністю за допомогою нечіткої логіки.	Обмежене впровадження в прогнозуванні електроенергії порівняно з іншими методами.	Помірна	Висока	Помірна	Може підійти для конкретних сценаріїв.
17			Пропонує інтерпретацію обґрунтування прогнозів.					
18	Expert System	На основі знань	Захоплює досвід і знання людини для прийняття обґрунтованих рішень.	Обмежена можливість самостійного прогнозування.	Обмежена	Висока	Це займає багато часу і потребує досвіду.	Надає цінну інформацію щодо невизначеності.
19			Добре інтегрується з іншими методами прогнозування.	Отримання знань може бути трудомістким і дорогим.				
20	Ансамблеве прогнозування	Гібридний	Поєднує в собі кілька моделей для надійності та підвищеної точності.	Вимагає досвіду у виборі та комбінуванні різних методів.	Змінюється в залежності	Змінюється в залежності	Змінюється в залежності	Може використовувати і сильні сторони окремих методів.
21	Хольта - Вінтерса	Статистичний	Вловлює тенденції та сезонність.	Складніше, ніж експоненціальне згладжування.	Помірна	Помірна	Помірна	Може бути придатним із відповідними техніками.
22			Добре підходить для даних із тенденціями та сезонними моделями.	Може бути не ідеальним для дуже нелінійних зв'язків.				

### Рисунок 3.2 – Порівняльна таблиця методів прогнозування (2)

Отже, підводячи підсумок всього вище сказаного, можна зробити такий висновок, що статистичні моделі - це традиційні засоби прогнозування, які експлуатують історичні результати та статистичні методи під цю задачу. Методи МН ширше використовують різні алгоритми, які навчаються на даних, у тому числі такі популярні варіанти, як штучні нейронні мережі для фіксації складних взаємозв'язків. Економетричні моделі включають економічні теорії та дані, щоб зрозуміти, як ширші економічні тенденції впливають на попит на електроенергію. Ці категорії іноді можуть збігатися. Наприклад, деякі статистичні методи можна використовувати в рамках машинного навчання. Однак розуміння цих основних категорій забезпечує хорошу відправну точку для вивчення конкретних методів прогнозування, придатних для споживання електроенергії підприємством в умовах невизначеності.

### 3.2 Обраний метод

Для цієї задачі було обрано метод лінійної регресії. Він залишається цінним інструментом у прогнозуванні споживання електроенергії з кількох причин. По-перше, це допомагає визначити взаємозв'язки між різними факторами, такими як погода, рівень виробництва, час доби та споживання електроенергії. Це розуміння має вирішальне значення для розробки функцій у більш складних моделях прогнозування. Крім того, регресійні моделі пропонують високий рівень інтерпретації, полегшуючи розуміння та передачу зв'язків між прогнозними показниками та споживанням ЕЕ. Ця прозорість може бути цінним аргументом в даній темі.

В методі лінійної регресії модель вчиться на основі наявних даних. Процес навчання полягає у знаходженні оптимальних значень параметрів моделі, які найкращим чином апроксимують залежність між вхідними та вихідними змінними, і аналізує їх таким чином, щоб мінімізувати помилку прогнозування. Зазвичай процес навчання лінійної регресії включає наступні кроки. Спочатку проводиться підготовка

даних, що включає очищення їх від аномальних значень та пропусків, масштабування відомостей, можливе створення нових ознак з наявних та розбиття даних на тренувальний та тестовий набори [41]. Потім визначається функція втрат, яка визначає, наскільки добре модель відповідає даним. Після цього вибирається алгоритм оптимізації, який використовується для знаходження оптимальних значень параметрів моделі. Далі відбувається навчання моделі, процес, за якого модель адаптується до тренувальних даних, шукаючи оптимальні значення параметрів. Нарешті, після навчання моделі її ефективність оцінюється на тестовому наборі даних, щоб переконатися, що вона добре працює на нових даних.

Коли справа доходить до написання коду на Python для прогнозування споживання електроенергії, включення даної технології може бути корисним. Бібліотеки цієї мови, такі як Statsmodels, надають комплексні інструменти завдяки чому реалізація регресійних моделей є відносно простою. Крім того аналіз можна легко інтегрувати в сценарії або блокноти додатку разом з іншими методами прогнозування, дозволяючи дослідникам поєднувати сильні сторони регресії з більш просунутими методами, такими як алгоритми машинного навчання. Ця мова програмування пропонує надійні бібліотеки [43] візуалізації, як Matplotlib і Seaborn, котрі можна використовувати для показу результатів регресійного аналізу, допомагаючи зрозуміти взаємозв'язки між прогнозними показниками та споживанням ЕЕ і передавати результати зацікавленим сторонам.

### **3.3 Мова програмування Python**

Python — це високорівнева інтерпретована мова програмування загального призначення, створена Гвідо ван Россумом і вперше випущена в 1991 році. Він розроблений з акцентом на читабельність коду, а її синтаксис дозволяє програмістам виражати концепції меншою кількістю рядків коду, ніж це можливо в таких мовах, як C++ або Java [43].

Сама програма виконується особистим інтерпретатором рядок за рядком, що відрізняється від скомпільованих мов, де код спочатку перетворюється на специфічні для машини інструкції (рис. 3.3).



Рисунок 3.3 – Спрощена блок-схема роботи Python

Для створення програми, необхідно почати з написання коду. В цьому на допомогу приходять різні редактори (Visual Studio Code, PyCharm, Sublime Text, GNU Emacs), котрі мають обширні можливості: вказівки на помилки, допомога в написанні, інструменти налагодження тощо. Сам же код буде мати в собі операторів, функції і класи, які визначають логіку та функціональність. Зберігання надає додатку розширення .py і буде мати вигляд сценарію Python, яке діє як набір інструкцій до нього. Після запуску в гру вступає інтерпретатор, котрий зчитує інформацію, перекладає його у формат зрозумілий комп'ютеру та виконує все те, що було визначено в додатку. З попередньо встановленою бібліотекою, яка поставляється з цією мовою програмування, надається і багатий набір вбудованих функцій, модулів і структур даних для типових завдань. В даній збірці охоплюється обробка файлів, мережа та маніпулювання даними, що дозволяє безпосередньо використовувати ці функції без необхідності додаткового встановлення. Окрім вбудованого пакету існує безліч сторонніх бібліотек інших розробників. Кожна із них пропонує спеціалізовані функції для своєї галузі.

Python відомий своїм простим і зрозумілим синтаксисом з акцентом на зручності читання, яка стає ідеальною мовою як для початківців, так і для досвідчених програмістів. Провівши дослідження на початку 2024 року серед більш як дев'яти тисяч українських ІТ спеціалістів, було виявлено, що дана мова є третьою по числу використання (рис. 3.4). Це кросплатформна програма і може працювати на різних операційних системах, таких як Windows, Linux, macOS тощо, а саме головне без жодних змін у коді. Стандартна бібліотека є обширною та універсальною, надає широкий спектр модулів і функцій для таких завдань, як файловий ввід/вивід, мережеве спілкування, обробка даних тощо. Має багату та постійно зростаючу екосистему з величезною кількістю сторонніх бібліотек і фреймворків: NumPy, pandas, TensorFlow, Flask і Django [44]. Пакети пов'язані з математикою, фізикою та іншими науковими галузями (NumPy, pandas і Matplotlib) роблять цей додаток більш використовуваним для наукових досліджень/обчислень, аналізу даних, обробки та візуалізації, . Також є незамінною та провідною мовою в галузі машинного навчання та штучного інтелекту з такими збірками, як TensorFlow, Keras і scikit-learn. Python є динамічно типізованою мовою, що означає, що тип змінної визначається під час виконання, а не під час компіляції. Це забезпечує більшу гнучкість коду, але також може призвести до помилок під час виконання, якщо не оброблятися належним чином.

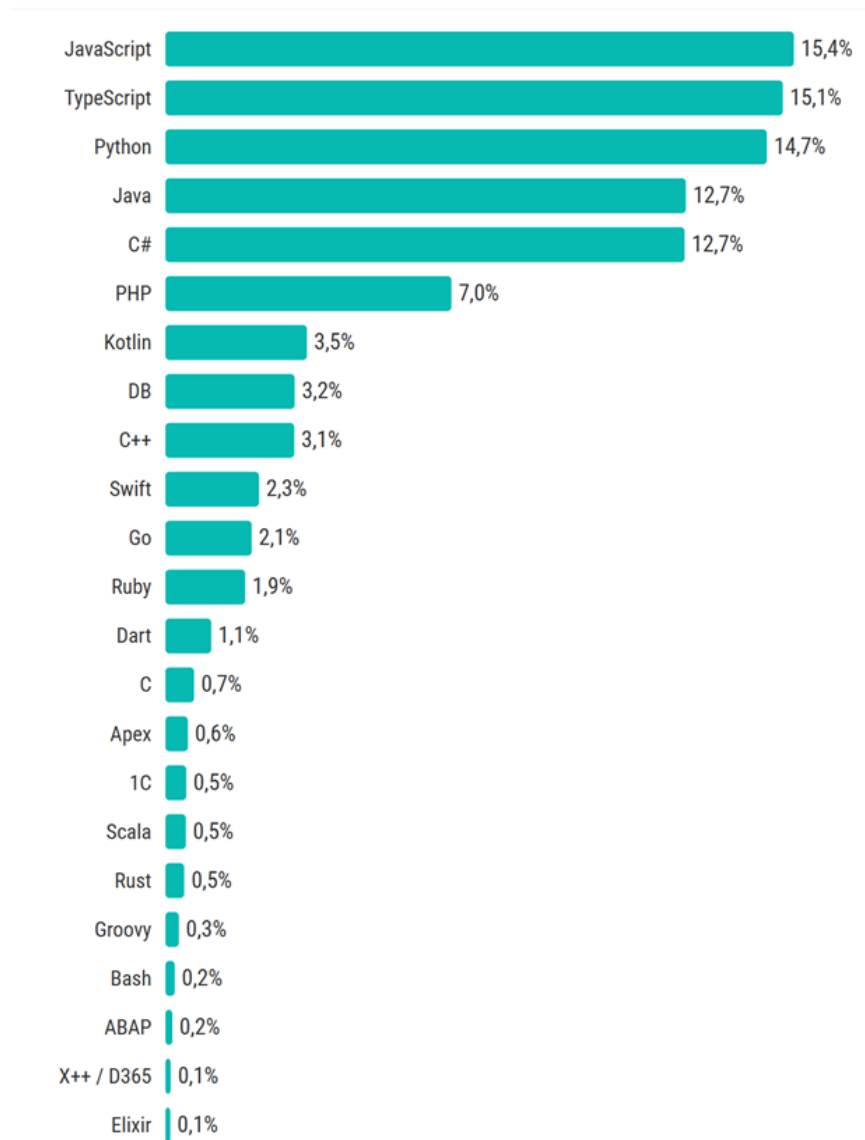


Рисунок 3.4 – Рейтинг мов програмування

На даний час Python широко використовується в різних областях. Його простота та наявність потужних фреймворків роблять його популярним вибором для розробки. Його простота, легкість використання та велика стандартна бібліотека роблять його чудовим вибором для автоматизації завдань і створення сценаріїв.

#### 4. СТВОРЕННЯ ТА ОПИСАННЯ ПРОГРАМИ

У цьому розділі детально описано розробку програми Python для прогнозування споживання електроенергії підприємством в умовах невизначеності. Програма використовує обраний метод і використовує можливості мови програмування для обробки даних, побудови моделей і прогнозування.

Для виконання роботи задіялись 2-а excel-файли з замірами (рисунок 4.1 – 4.2). В них знаходяться часи замірів з кроком 30 хв. для того, щоб зрозуміти в який час використовуються максимальна кількість ЕЕ. Також номер заміру, скільки енергії було використано та показання.

	A	B	C	D	E	F
1	Date	Time	Number	Energy	Indications	
2	01.01.2012	0:30:00	1	25	1668	
3	01.01.2012	1:00:00	2	27	1668	
4	01.01.2012	1:30:00	3	30	1669	
5	01.01.2012	2:00:00	4	30	1669	
6	01.01.2012	2:30:00	5	27	1669	
7	01.01.2012	3:00:00	6	24	1669	
8	01.01.2012	3:30:00	7	25	1669	
9	01.01.2012	4:00:00	8	25	1669	
0	01.01.2012	4:30:00	9	25	1669	
1	01.01.2012	5:00:00	10	24	1669	
2	01.01.2012	5:30:00	11	25	1669	
3	01.01.2012	6:00:00	12	25	1670	
4	01.01.2012	6:30:00	13	25	1670	
5	01.01.2012	7:00:00	14	25	1670	
6	01.01.2012	7:30:00	15	23	1670	
7	01.01.2012	8:00:00	16	17	1670	
8	01.01.2012	8:30:00	17	16	1670	
9	01.01.2012	9:00:00	18	15	1670	
0	01.01.2012	9:30:00	19	16	1670	
1	01.01.2012	10:00:00	20	15	1670	
2	01.01.2012	10:30:00	21	16	1670	
3	01.01.2012	11:00:00	22	16	1670	
4	01.01.2012	11:30:00	23	16	1670	
5	01.01.2012	12:00:00	24	15	1670	
6	01.01.2012	12:30:00	25	18	1670	
7	01.01.2012	13:00:00	26	16	1671	
8	01.01.2012	13:30:00	27	17	1671	
9	01.01.2012	14:00:00	28	16	1671	
0	01.01.2012	14:30:00	29	17	1671	
1	01.01.2012	15:00:00	30	18	1671	
2	01.01.2012	15:30:00	31	17	1671	
3	01.01.2012	16:00:00	32	17	1671	

Рисунок 4.1 – заміри енергетичної установи за 2012 рік.



	A	B	C	D
1	Date	Time	Number	Energy
2	01.01.2013	0:30:00	1	46
3	01.01.2013	1:00:00	2	47
4	01.01.2013	1:30:00	3	47
5	01.01.2013	2:00:00	4	45
6	01.01.2013	2:30:00	5	46
7	01.01.2013	3:00:00	6	47
8	01.01.2013	3:30:00	7	47
9	01.01.2013	4:00:00	8	47
10	01.01.2013	4:30:00	9	45
11	01.01.2013	5:00:00	10	46
12	01.01.2013	5:30:00	11	47
13	01.01.2013	6:00:00	12	48
14	01.01.2013	6:30:00	13	48
15	01.01.2013	7:00:00	14	45
16	01.01.2013	7:30:00	15	38
17	01.01.2013	8:00:00	16	34
18	01.01.2013	8:30:00	17	34
19	01.01.2013	9:00:00	18	36
20	01.01.2013	9:30:00	19	36
21	01.01.2013	10:00:00	20	34
22	01.01.2013	10:30:00	21	36
23	01.01.2013	11:00:00	22	36
24	01.01.2013	11:30:00	23	35
25	01.01.2013	12:00:00	24	33
26	01.01.2013	12:30:00	25	38
27	01.01.2013	13:00:00	26	39
28	01.01.2013	13:30:00	27	36
29	01.01.2013	14:00:00	28	36
30	01.01.2013	14:30:00	29	34
31	01.01.2013	15:00:00	30	38
32	01.01.2013	15:30:00	31	36

Рисунок 4.2 – заміри 2013 року.

Бібліотеки в мовах програмування — це важливі колекції попередньо написаних модулів коду, які надають багаторазові функції, класи та утиліти для спрощення та прискорення розробки програмного забезпечення. Крім того, вони часто проходять суворе тестування та оптимізацію, підвищуючи надійність і продуктивність у різноманітних обчислювальних середовищах. Для початку необхідно написати про критичні пакети цієї роботи. Хоча список модулів в програмі великий, все таки ось перелік головних бібліотек.

**Pandas** - ця бібліотека діє як основа для маніпулювання даними. Це дозволяє легко обробляти історичні дані про споживання електроенергії. Вона чудово завантажує дані з різних джерел (CSV, Excel), очищає їх (обробляє пропущені значення, викиди) і готує до аналізу. Крім того, допомагає розділити відомості на

набори для навчання та тестування, що має вирішальне значення для оцінки моделі прогнозування.

**NumPy** надає програмі потужні чисельні обчислювальні можливості. Це дозволяє виконувати складні обчислення, агрегації та інші математичні операції з документами про споживання електроенергії. Ефективні структури даних, які називаються масивами NumPy, ідеально підходять для цих обчислень. Крім того, операції лінійної алгебри можуть бути необхідними для певних моделей прогнозування, зокрема статистичних моделей.

**Scikit-learn.** Вона єдина бібліотека для машинного навчання. Якщо обираєте метод прогнозування на основі машинного навчання, scikit-learn надає величезну колекцію інструментів і алгоритмів для побудови та оцінки вашої моделі. Можна використовувати його для впровадження різних алгоритмів регресії (наприклад, лінійної регресії або випадкового лісу), навчання та тестування моделі на історичних даних і навіть тонкого налаштування її параметрів для оптимальної продуктивності.

**Matplotlib,** незважаючи на те, що він не важливий для основних функцій програми, це потужний інструмент для створення візуалізацій. Це дозволяє створювати статичні, анімовані та навіть інтерактивні візуалізації результату. Це може бути надзвичайно корисним для таких завдань, як: візуалізація історичних моделей споживання електроенергії (через графіки часових рядів), дослідження взаємозв'язків між змінними, щоб допомогти у виборі моделі, і оцінка ефективності моделі шляхом порівняння прогнозованих і фактичних значень.

**SciPy.** Створений на основі NumPy, SciPy пропонує додаткові функції для наукових обчислень. Залежно від обраного методу прогнозування, він може бути корисним для виконання статистичного аналізу (перевірки гіпотез), застосування процедур оптимізації під час навчання моделі та використання спеціалізованих математичних функцій, пов'язаних із конкретними методами прогнозування.

Програма використовує лінійну регресію як основний метод прогнозування споживання електроенергії. Він є відносно простим для реалізації та розуміння, ефективним, з точки зору обчислень, що робить його придатним для обробки великих наборів даних. Можливість зрозуміти, як різні фактори впливають на споживання

електроенергії із-за того, що коефіцієнти в лінійному рівнянні представляють вплив кожної незалежної змінної (час, сезон, день тижня тощо) на залежну змінну (те, що необхідно передбачити – у цьому випадку споживання електроенергії). Додаток дотримується структурованого підходу, розділеного на кілька функцій для підготовки даних, навчання моделі, прогнозування та оцінки помилок. Програма включає етапи попередньої обробки відомостей для обробки відсутніх значень, перетворення типів даних і створення нових функцій (наприклад, день тижня - `get_day`, сезон - `get_season`).

По-перше, програма починається із підготовки даних. Окремі функції обробляють завантаження навчальних і тестових даних, їх очищення та перетворення у відповідний формат для моделі. Це включає такі завдання, як розподіл дат на місяць і сезон, нормалізація значень часу та створення нових функцій на основі інформації про час. Після підготовки вступає в дію побудова моделі. Додаток починається з випадкової ініціалізації коефіцієнтів для моделі лінійної регресії. Ці коефіцієнти по суті представляють ваги, присвоєні різним факторам, що впливають на споживання електроенергії.

```
def proceed_train_data(file_path):
    # Зчитуємо дані з файлу
    data = pd.read_csv(file_path, sep = ';')
    # Розділяємо стовпчик 'місяць/день/рік' на окремі стовпці 'місяць' та 'пора року'
    print(data)
    data["month"] = data.iloc[:, 0].str.split('.', expand=True)[0]
    data["season"] = data["month"].astype(int).apply(get_season)
    # Нормалізуємо значення в стовпці 'час' від 0 до 1
    print(data.iloc[2])
    data["time"] = data["Time"].apply(lambda x: (int(x.split(':')[0]) * 60 + int(x.split(':')[1])) / (24 * 60))
    # Додаємо стовпчик 'день неділі', де 1 відповідає неділі, а інші дні тижня - 0
    data["Weekday"] = (pd.to_datetime(data["Date"], format='%d.%m.%Y').dt.dayofweek == 6).astype(int)
    data.rename(columns={'Energy': 'value'}, inplace=True)
    # Додаємо стовпці 'день' та 'ніч' зі значеннями 1 та 0 відповідно
    data["Day"] = data["Time"].apply(get_day)
    data["Night"] = data["Time"].apply(get_night)
    return data

def process_test_data(test_data):
    # Розділяємо перший стовпчик на місяць та пору року
    test_data["month"] = test_data.iloc[:, 0].str.split('.', expand=True)[0]
    print(test_data["month"])
    test_data["season"] = test_data["month"].astype(int).apply(get_season)
    # Робимо третій стовпчик часом і нормалізуємо його
    test_data["time"] = test_data["Time"].apply(lambda x: (int(str(x).split(':')[0]) * 60 + int(str(x).split(':')[1])) / 30 / 48)
    # Записуємо сьомий стовпчик як день неділі
    test_data["Weekday"] = pd.to_datetime(test_data["Date"], format='%d.%m.%Y').dt.dayofweek + 1
    # Переміщуємо значення в четвертий стовпчик
    # test_data.rename(columns={'value': '4'}, inplace=True)
    # test_data.rename(columns={'Energy': 'value'}, inplace=True)

    # Перетворюємо значення 'день' і 'ніч'
    test_data["Day"] = test_data["Time"].apply(get_day)
    test_data["Night"] = test_data["Time"].apply(get_night)
    # Повертаємо оброблені дані
    return test_data
```

Рисунок 4.3 - функції `proceed_train_data` і `process_test_data`

Тепер настає основний тренувальний процес. Тут програма ітераційно уточнює спочатку вгадані коефіцієнти. У кожній ітерації він обчислює прогнозовані значення споживання на основі поточних коефіцієнтів і різних характеристик, таких як місяць, день тижня та сезон. Щоб включити поточний час, програма враховує лише відповідний день тижня та сезон під час прогнозування. Порівнюючи прогнозовані значення з фактичними значеннями споживання, додаток обчислює метрику помилки (середньоквадратичну помилку), щоб оцінити, наскільки добре модель відповідає даним навчання. Озброївшись цією інформацією про помилку, вона коригує коефіцієнти за допомогою техніки, що називається градієнтним спуском. Це, по суті, змінює фактор в напрямку, який мінімізує похибку між прогнозованим і фактичним споживанням. Щоб запобігти надмірній залежності моделі від конкретних шаблонів тренувальних даних (переобладнання), програма також включає техніку, яка називається L1 регуляризацією під час навчання.

```
def train_model(data, coefficients, learning_rate, regularization_lambda, max_iterations, tolerance, train_data):
    iteration = 0
    prev_mse = float('inf')

    while True:
        # Обчислюємо регресію з поточними коефіцієнтами
        regression_values = regression[data, coefficients]
        # Обчислюємо середньоквадратичну помилку
        mse = mean_squared_error(regression_values, data['value'])
        # Перевіряємо точність
        if abs(mse - prev_mse) <= tolerance:
            break
        # Перевірка на перенавчання за допомогою L1 регуляризації
        if iteration > 0:
            regularization_term = regularization_lambda * np.sign(coefficients)
            print("+++++++")
            print(data.iloc[:, 3:])
            print("+++++++")
            print(regression_values - data['value'])
            coefficients -= learning_rate * (np.dot(regression_values - data['value'], data.iloc[:, 3]) / len(data) + regularization_term)
        else:
            # Оновлюємо коефіцієнти без регуляризації на першій ітерації
            coefficients = update_coefficients(coefficients, regression_values, data['value'], learning_rate, train_data)
        # Виводимо значення помилки кожні 20 ітерацій
        if iteration % 20 == 0:
            print(f"Iteration {iteration}: MSE = {mse}")
            # Питаємо, чи продовжити навчання
            answer = input("Continue training? (yes/no): ")
            if answer.lower() != 'yes':
                break
        # Оновлюємо значення помилки з попередньої ітерації
        prev_mse = mse
        iteration += 1

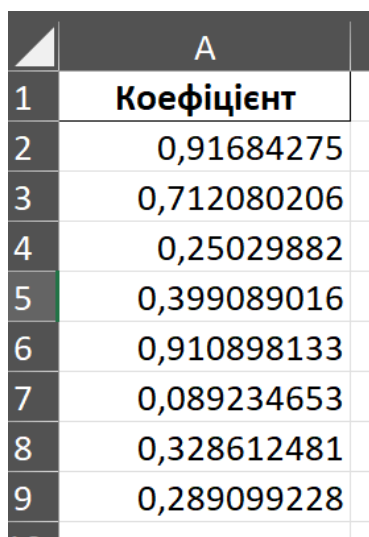
    return coefficients
```

Рисунок 4.4 - Функція train\_model

Після завершення процесу навчання вона має навчену модель з оптимізованими коефіцієнтами. Ця модель по суті представляє вивчений зв'язок між споживанням електроенергії та різними факторами впливу. Тепер настав час використовувати цю модель за прямим призначенням: прогноз. Програма використовує навчену модель для прогнозування значень споживання електроенергії для невидимих тестових даних. Він застосовує ту саму логіку, що й під час навчання, але цього разу на тестових даних для створення прогнозів споживання.

Нарешті, щоб оцінити продуктивність моделі на основі тестових даних, програма обчислює середню квадратичну похибку між прогнозованим і фактичним значеннями споживання. Крім того, він генерує візуальне представлення, відображаючи прогнозовані значення проти фактичних значень, дозволяючи отримати чіткішу картину того, наскільки прогнози моделі відповідають реальним моделям споживання.

В кінці створюються два файли excel в папці з проектом: coefficients.xlsx (рисунок 4.5) та прогнозовані\_значення.xlsx (рисунок 4.6).



	A
1	<b>Коефіцієнт</b>
2	0,91684275
3	0,712080206
4	0,25029882
5	0,399089016
6	0,910898133
7	0,089234653
8	0,328612481
9	0,289099228

Рисунок 4.5 – файл coefficients.xlsx та дані в ньому

	A	B	C	D	E	F	G	H	I	J	K
1	Date	Time	Number	Energy	month	season	time	Weekday	Day	Night	прогнозоване значення
2	01.01.2013	0:30:00	1	46 01		1	0,020833	2	0	0	0
3	01.01.2013	1:00:00	2	47 01		1	0,041667	2	0	0	0
4	01.01.2013	1:30:00	3	47 01		1	0,0625	2	0	0	0
5	01.01.2013	2:00:00	4	45 01		1	0,083333	2	0	0	0
6	01.01.2013	2:30:00	5	46 01		1	0,104167	2	0	0	0
7	01.01.2013	3:00:00	6	47 01		1	0,125	2	0	0	0
8	01.01.2013	3:30:00	7	47 01		1	0,145833	2	0	0	0
9	01.01.2013	4:00:00	8	47 01		1	0,166667	2	0	0	0
10	01.01.2013	4:30:00	9	45 01		1	0,1875	2	0	0	0
11	01.01.2013	5:00:00	10	46 01		1	0,208333	2	0	0	0
12	01.01.2013	5:30:00	11	47 01		1	0,229167	2	0	0	0
13	01.01.2013	6:00:00	12	48 01		1	0,25	2	0	0	0
14	01.01.2013	6:30:00	13	48 01		1	0,270833	2	0	0	0
15	01.01.2013	7:00:00	14	45 01		1	0,291667	2	0	0	0
16	01.01.2013	7:30:00	15	38 01		1	0,3125	2	0	0	0
17	01.01.2013	8:00:00	16	34 01		1	0,333333	2	0	0	0
18	01.01.2013	8:30:00	17	34 01		1	0,354167	2	0	0	0
19	01.01.2013	9:00:00	18	36 01		1	0,375	2	0	0	0
20	01.01.2013	9:30:00	19	36 01		1	0,395833	2	0	0	0
21	01.01.2013	10:00:00	20	34 01		1	0,416667	2	0	0	0
22	01.01.2013	10:30:00	21	36 01		1	0,4375	2	0	0	0
23	01.01.2013	11:00:00	22	36 01		1	0,458333	2	0	0	0
24	01.01.2013	11:30:00	23	35 01		1	0,479167	2	0	0	0
25	01.01.2013	12:00:00	24	33 01		1	0,5	2	0	0	0
26	01.01.2013	12:30:00	25	38 01		1	0,520833	2	0	0	0

Рисунок 4.6 – створений файл прогнозовані\_значення.xlsx та результати виконання

Web-додаток представляє собою маленьке вікно (рисунок 4.7), в якому необхідно буде додати 2 файли, натиснути спочатку кнопку «<>», а вже потім «>». Після чого, програма надасть графік прогнозу (рисунок 4.8), де можна подивитися результати та оцінити значення виміру.

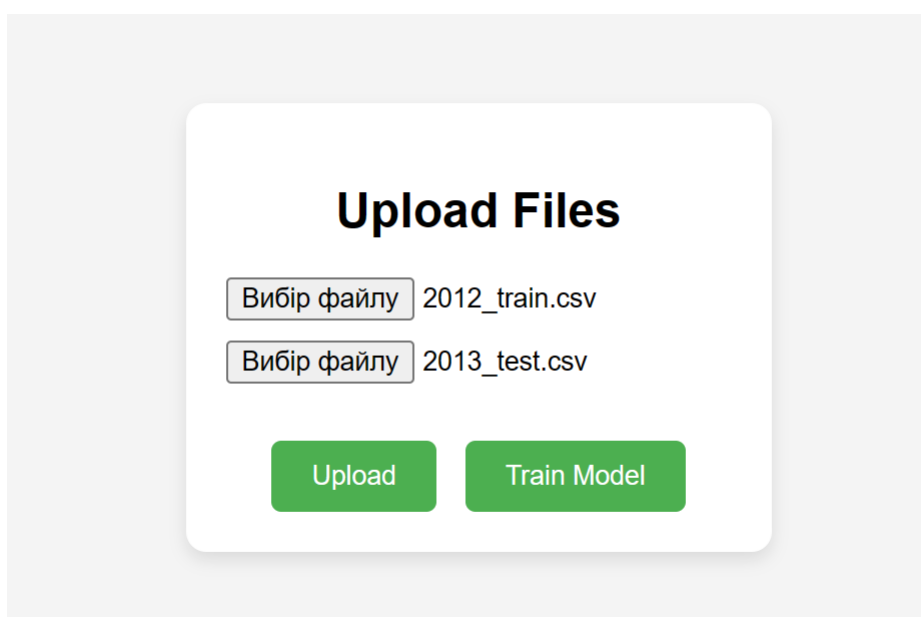


Рисунок 4.7 – Головне вікно програми

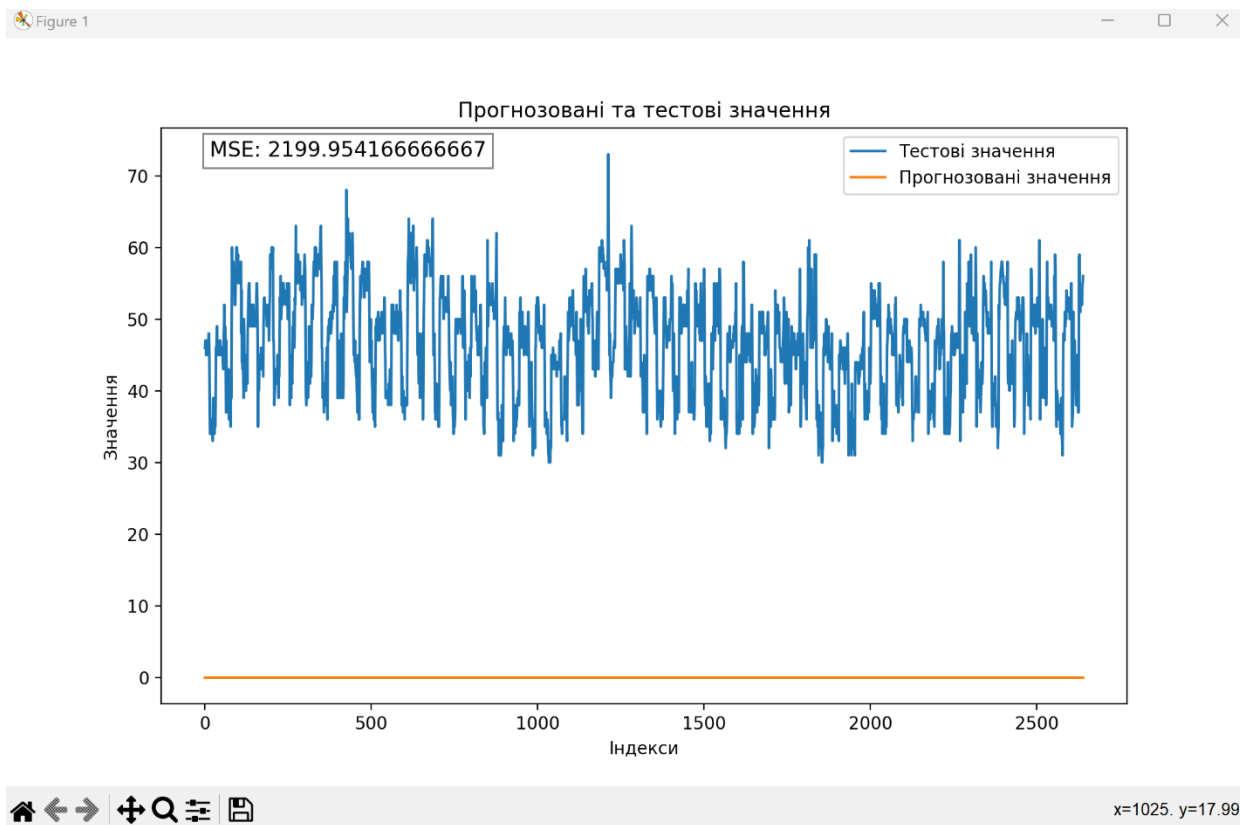


Рисунок 4.8 – Результат виконання програми

## ВИСНОВОК

В роботі досліджено різні підходи до прогнозування енергоспоживання, генерації та вартості в контексті ринку електроенергії. Спостереження підкреслило важливість точного передбачення в енергетичному секторі, оскільки це дозволяє підприємствам оптимізувати використання ресурсів, зменшити витрати та покращити стійкість.

Проведений аналіз використання штучного інтелекту: як він може допомогти з розв'язком даної проблеми, що нового принесе з собою у цю сферу, перераховування його потенційних сильних сторін тощо.

Крім того, в роботі проаналізовано існуючі методи прогнозування електроспоживання, такі як метод найближчого сусіда, нечіткого регресійного аналізу, експоненціального згладжування, а також метод авторегресії та ковзного середнього. Вони можуть бути вжиті для прогнозу споживання енергії в різних контекстах і надання підприємствам цінного розуміння їх використання.

Таким чином, теми, розглянуті в роботі, дають цінне уявлення про існуючі методи передбачення споживання електроенергії. Застосування цих підходів зменшить витрати підприємства на ЕЕ, поліпшить свою стійкість і залишиться конкурентоспроможним на ринку, що швидко розвивається.

Розроблений додаток продемонстрував добре структурований робочий процес з окремими функціями, присвяченими підготовці даних, навчання моделі, прогнозуванню та оцінці. Підготовка даних забезпечила узгодженість і вирішила потенційні проблеми, як-от відсутні значення та нестандартні формати.

Хоча основна функціональність покладалася на лінійну регресію, програма визнала її обмеження, зокрема припущення про лінійний зв'язок між змінними.

Додаток забезпечує міцну основу для прогнозування споживання електроенергії за допомогою лінійної регресії. З подальшим удосконаленням і дослідженням альтернативних методів це може стати цінним інструментом для підприємств, щоб ефективніше керувати своїми потребами в енергії.



## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Прогнозування споживання електричної енергії електротехнічних комплексів міської електричної мережі. 2022. [Електронний ресурс] Режим доступу до статистичних даних: <https://isg-journal.com/isjea/article/view/23/j-isjea-20220103.13>
2. Енергетичний баланс України за 2024 рік [Електронний ресурс] // Державна служба статистики України. Режим доступу до статистичних даних: [www.ukrstat.gov.ua](http://www.ukrstat.gov.ua)
3. Прогнозування показників енергоспоживання, генерації і вартості отриманої енергії. 2020 рік [Електронний ресурс] Режим доступу до статистичних даних: [https://www.tech.vernadskyjournals.in.ua/journals/2020/2\\_2020/part\\_1/40.pdf](https://www.tech.vernadskyjournals.in.ua/journals/2020/2_2020/part_1/40.pdf)
4. Глібка Н. Л. Прогнозування споживання електричної енергії промисловим підприємством: дис. маг. 2024 / Глібка Н. Л. Київ, 2024, 88 с.
5. Симець І. І. Моделі і методи прогнозування та аналізу надійності технічних систем з урахуванням процесу розробки ПЗ: дис. док. філ. 2022 / Симець І.І. Львів, 2022, 188 с.
6. Яковуна, V., Seniv, M., Symets, I. The Relation between Software Development Methodologies and Factors Affecting Software Reliability // Proceedings of IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), CSIT 2020, 23-26 Sept. 2020. - Zbarazh, Ukraine. - 377 - 381; <https://ieeexplore.ieee.org/document/9321937>.
7. Тимчук С. О. Прогнозування електроспоживання на основі нечіткого регресійного аналізу / С. О. Тимчук, І. А. Катюха // Вісник Харківського національного технічного університету сільського господарства імені Петра Василенка. Проблеми енергозабезпечення та енергозбереження в АПК України. – Х.: ХНТУСГ, 2012. – Вип. 130. – С. 36-38.
8. Катюха І. А. Прогнозні моделі електричних навантажень розподільчих мереж в умовах невизначеності вихідної інформації: дис. канд. тех. наук. 14.07.2017 / Катюха І. А. Мелітополь, 2017, 150 с.

9. Pooley John. Quick Start Guide to Energy Monitoring & Targeting (M&T) [Електронний ресурс] Режим доступу: <https://www.oursouthwest.com/eemg/notices/effective-energy-mgt-mandtguide.pdf>
10. Directive (eu) 2019/944 of the european parliament and of the council. 2022. [Електронний ресурс] Режим доступу до статистичних даних: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L\\_.2019.158.01.0125.01.ENG&toc=OJ:L:2019:158:TOC](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.158.01.0125.01.ENG&toc=OJ:L:2019:158:TOC)
11. Guidance and recommendation on energy storage 2021. [Електронний ресурс] Режим доступу до статистичних даних: [https://energy.ec.europa.eu/topics/research-and-technology/energy-storage\\_en](https://energy.ec.europa.eu/topics/research-and-technology/energy-storage_en)
12. Regulation (eu) no 1227/2011 of the european parliament and of the council: on wholesale energy market integrity and transparency 2011. [Електронний ресурс] Режим доступу до статистичних даних: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32011R1227>
13. Успішний старт: перший крок роботи Оптового ринку електричної енергії в Україні. 2020. [Електронний ресурс] Режим доступу до статистичних даних: <https://energysecurityua.org/ua/novyny/uspishnii-start-pershii-rik-roboty-optovogo-rynku-elektrychnoyi-enerhiyi-v-ukrayini/>
14. Технології ШІ проти небалансів: сучасна система прогнозування виробництва електроенергії з ВДЕ. 2023. [Електронний ресурс] Режим доступу до статистичних даних: <https://kness.energy/news/tehnologii-shi-proti-nebalansiv-suchasna-sistema-prognozuvannya-virobnictva-elektroenergii-z-vde/>
15. Про схвалення Концепції розвитку штучного інтелекту в Україні: Розпорядження Кабінету Міністрів України від 02.12.2020. № 1556-р. URL: <https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#Text>
16. OECD Framework for the Classification of AI Systems. OECD Digital Economy Papers. February 2022. № 323. P. 80.
17. Суходоля О. М. Штучний інтелект в енергетиці: аналітична доповідь д-р наук з держ. упр., професор. 2022 / Суходоля О. М. Київ, 2022, 49 с. [Електронний

- ресурс] Режим доступу до статистичних даних:  
[https://www.researchgate.net/publication/361747571\\_Artificial\\_Intelligence\\_in\\_Energy](https://www.researchgate.net/publication/361747571_Artificial_Intelligence_in_Energy)
18. Electricity consumption prediction using machine learning. 2023. [Електронний ресурс] Режим доступу до статистичних даних: [https://www.e3s-conferences.org/articles/e3sconf/pdf/2023/28/e3sconf\\_icmed-icmpe2023\\_01048.pdf](https://www.e3s-conferences.org/articles/e3sconf/pdf/2023/28/e3sconf_icmed-icmpe2023_01048.pdf)
19. Forecasting Energy Consumption using Machine Learning and AI. 2024. . [Електронний ресурс] Режим доступу до статистичних даних: <https://www.dexma.com/blog-en/forecasting-energy-consumption-using-machine-learning-and-ai/>
20. What's new in a-Gnostics 2.0. Industrial AI service focused on anomaly detection and equipment failure prediction. 2021. [Електронний ресурс] Режим доступу до статистичних даних: [What's new in a-Gnostics 2.0. Industrial AI service focused on anomaly detection and equipment failure prediction | by a-Gnostics | Medium](https://medium.com/a-gnostics/whats-new-in-a-gnostics-2-0-industrial-ai-service-focused-on-anomaly-detection-and-equipment-failure-prediction)
21. Україна стала членом "енергетичного Євросоюзу" під час війни. Що це означає?. 2022. [Електронний ресурс] Режим доступу до статистичних даних: <https://www.bbc.com/ukrainian/news-60764933>
22. Мокін О. Б. Метод ідентифікації моделі авторегресії-ковзного середнього АРКС( $p, q$ ) з довільними значеннями порядків  $p, q$ , який узагальнює методику Юла – Уокера. / О. Б. Мокін, В. Б. Мокін, Б. І. Мокін. // Наукові праці Вінницького національного технічного університету. - 2014. - Вип. 2. – Режим доступу <https://praci.vntu.edu.ua/index.php/praci/article/view/406/404>
23. Бокс Дж. Анализ временных рядов. Прогноз и управление. Вып. 1, Вып. 2 / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1974. – 408 с.
24. Мокін Б. І. Математичні методи ідентифікації динамічних систем / Б. І. Мокін, В. Б. Мокін, О. Б. Мокін. – Вінниця: ВНТУ, 2010. – 260 с.
25. Застосування експоненціального згладжування для вимірювання поточних координат об'єктів управління. 2014 [Електронний ресурс] Режим доступу до статистичних даних: <https://ksua-news.ztu.edu.ua/wp-content/uploads/2014/11/lrpsy10.pdf>

26. Принцип експоненціального згладжування, як засіб мінімізації при енергозберігаючих заходах. 2017. [Електронний ресурс] Режим доступу до статистичних даних: <https://ena.lpnu.ua:8443/server/api/core/bitstreams/c1a85909-498c-4b09-b80d-279a55035155/content>
27. Братусь О. В. Метод двобічного експоненційного згладжування для відновлення динамічних процесів / О. В. Братусь, В. М. Подладчиков, П. І. Бідюк // Наукові вісті НТУУ «КПІ» : міжнародний науково-технічний журнал. – 2016. – № 6(110). – С. 15–21.
28. Метод «найближчого сусіда» або системи міркувань на основі аналогічних випадків. 2019. [Електронний ресурс] Режим доступу до статистичних даних: <https://studfile.net/preview/7818687/page:2/>
29. Маннинг К.Д., Рагхаван П., Шютце Х. Введение в информационный поиск.— М.: Изд-во «Вильямс», 2011.
30. Teng Hu, Weina Niu, Xiaosong Zhang, Xiaolei Liu, Jiazhong Lu, Yuan Liu, “An Insider Threat Detection Approach Based on Mouse Dynamics and Deep Learning”, Security and Communication Networks, 2019
31. Теоретичні та практичні аспекти використання математичних методів та інформаційних технологій в освіті і науці. 2021. [Електронний ресурс] Режим доступу до статистичних даних: [https://kubg.edu.ua/images/stories/Departaments/nmc.nd/naukovi\\_temy/zvit\\_0116U004625.pdf](https://kubg.edu.ua/images/stories/Departaments/nmc.nd/naukovi_temy/zvit_0116U004625.pdf)
32. Катюха І. А. Прогнозування електроспоживання на основі нечіткого регресійного аналізу / С. О. Тимчук, І. А. Катюха // Проблеми енергозабезпечення та енергозбереження в АПК України: вісник ХНТУСГ, 2012. – Вип. 130. – С. 51–53.
33. J.J. Buckley, L.J. Jowers. Fuzzy Linear Regression I. Studies in Fuzziness and Soft Computing, volume 22, ISBN 978-3-540-76289-8. Springer, 2008. A.F. Shapiro. Fuzzy regression models. In <https://www.soa.org/globalassets/assets/files/static-pages/research/arch/2006/arch06v40n1-ii.pdf>

34. J. Kacprzyk, M. Fedrizzi (Ed.). Fuzzy Regression Analysis. Studies in Fuzziness and Soft Computing, ISBN-13: 978- 3790805918. Publisher: Physica-Verlag HD, 1992.
35. R.E. Moore. Methods and Applications of Interval Analysis. SIAM (Society for Industrial and Applied Mathematics) Philadelphia, 1979
36. Tanaka, H., Uejima, S. and Asai, K (1982) “Linear regression analysis with fuzzy model,” IEEE Transactions on Systems, Man and Cybernetics, 12(6), 903-907.
37. V. Novak. Fuzzy množiny a jejich aplikace (in Czech). Studies in Fuzziness and Soft Computing, ISBN 80-03-00325-3. SNTL Praha, 1990.
38. Teviashev A.D., Tkachenko V.F., Sizova N.D., Kostarev D.B. Information and analytical system for monitoring the energy resources of the enterprise // ICT-2020, присвяченої 90-річчю Харківського національного університету радіоелектроніки: збірник праць 9-ї Міжнародної науково-технічної конференції (18-20.11.2020 р.)
39. Top 5 technologies used in the energy sector 2023. [Електронний ресурс] Режим доступу до статистичних даних: <https://energy-oil-gas.com/news/top-5-technologies-used-in-energy-sector/>
40. Digital transformation in the energy industry: overview and tips 2024. [Електронний ресурс] Режим доступу до статистичних даних: <https://waverleysoftware.com/blog/digital-transformation-in-the-energy-industry/#:~:text=Smart%20Grids%3A%20Energy%20companies%20are,as%20wind%20and%20solar%20power.>
41. Єріна А. М. Статистичне моделювання та прогнозування: Навч. посібник. — К.: КНЕУ, 2001. — 170 с.
42. Методи та системи штучного інтелекту. 2016. [Електронний ресурс] Режим доступу до статистичних даних: <https://dspace.uzhnu.edu.ua/jspui/bitstream/lib/16306/1/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4%D0%B8%20%D1%82%D0%B0%20%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B8%20%D1%88%D1%82%D1%83%D1%87%D0%BD%D0%BE%D0%B3%D0%BE%20%D1%96%D0%BD%D1%82%D0%B5%D0%BB%D0%B5%D0%BA%D1%82%D1%83.%20%D0%9A%D0%>

[BE%D0%BD%D1%81%D0%BF%D0%B5%D0%BA%D1%82%20%D0%BB%D0%B5%D0%BA%D1%86%D1%96%D0%B9.pdf](#)

43. Python для обробки даних і моделювання фінансово-економічних показників. 2021. [Електронний ресурс] Режим доступу до статистичних даних: <https://itce.vntu.edu.ua/index.php/itce/article/view/823/544>
44. Рейтинг мов програмування 2024. TypeScript в трійці лідерів, Python з'являється у всіх нішах, а Rust — улюблена мова. 2024. [Електронний ресурс] Режим доступу до статистичних даних: <https://dou.ua/lenta/articles/language-rating-2024/>

## ДОДАТОК А

Програмний код **Diploma.py**

```
def proceed_train_data(file_path):
    # Зчитуємо дані з файлу
    data = pd.read_csv(file_path, sep = ';')
    # Розділяємо стовпець 'місяць/день/рік' на окремі стовпці 'місяць' та 'пора року'
    print(data)
    data["month"] = data.iloc[:, 0].str.split('.', expand=True)[0]
    data["season"] = data['month'].astype(int).apply(get_season)
    # Нормалізуємо значення в стовпці 'час' від 0 до 1
    print(data.iloc[2])
    data['time'] = data['Time'].apply(lambda x: (int(x.split(':')[0]) * 60 +
int(x.split(':')[1])) / (24 * 60))
    # Додаємо стовпець 'день неділі', де 1 відповідає неділі, а інші дні тижня - 0
    data['Weekday'] = (pd.to_datetime(data['Date'],
format='%d.%m.%Y').dt.dayofweek == 6).astype(int)
    data.rename(columns={'Energy': 'value'}, inplace=True)
    # Додаємо стовпці 'день' та 'ніч' зі значеннями 1 та 0 відповідно
    data['Day'] = data['Time'].apply(get_day)
    data['Night'] = data['Time'].apply(get_night)
    return data

def process_test_data(test_data):
    # Розділяємо перший стовпчик на місяць та пору року
    test_data['month'] = test_data.iloc[:, 0].str.split('.', expand=True)[0]
    print(test_data['month'])
    test_data["season"] = test_data['month'].astype(int).apply(get_season)
    # Робимо третій стовпчик часом і нормалізуємо його
    test_data['time'] = test_data['Time'].apply(lambda x: (int(str(x).split(':')[0]) * 60 +
int(str(x).split(':')[1])) / 30 / 48)
```

```

# Записуємо сьомий стовпчик як день неділі
test_data["Weekday"]
pd.to_datetime(test_data["Date"],format='%d.%m.%Y').dt.dayofweek + 1
# Переміщуємо значення в четвертий стовпчик
# test_data.rename(columns={'value': '4'}, inplace=True)
# test_data.rename(columns={'Energy': 'value'}, inplace=True)
# Перетворюємо значення 'день' і 'ніч'
test_data['Day'] = test_data['Time'].apply(get_day)
test_data['Night'] = test_data['Time'].apply(get_night)
# Повертаємо оброблені дані
return test_data

def train_model(data, coefficients, learning_rate, regularization_lambda,
max_iterations, tolerance, train_data):
    iteration = 0
    prev_mse = float('inf')
    while True:
        # Обчислюємо регресію з поточними коефіцієнтами
        regression_values = regression(data, coefficients)
        # Обчислюємо середньоквадратичну помилку
        mse = mean_squared_error(regression_values, data['value'])
        # Перевіряємо точність
        if abs(mse - prev_mse) <= tolerance:
            break
        # Перевірка на перенавчання за допомогою L1 регуляризації
        if iteration > 0:
            regularization_term = regularization_lambda * np.sign(coefficients)
            print(data.iloc[:, 3:])
            print(regression_values - data['value'])
            coefficients -= learning_rate * (np.dot(regression_values - data['value'],
data.iloc[:, 3]) / len(data) + regularization_term)

```



```

else:
    # Оновлюємо коефіцієнти без регуляризації на першій ітерації
    coefficients = update_coefficients(coefficients, regression_values,
data['value'], learning_rate, train_data)
    # Виводимо значення помилки кожні 20 ітерацій
    if iteration % 20 == 0:
        print(f'Iteration {iteration}: MSE = {mse}')
        # Питаємо, чи продовжити навчання
        answer = input("Continue training? (yes/no): ")
        if answer.lower() != 'yes':
            break
    # Оновлюємо значення помилки з попередньої ітерації
    prev_mse = mse
    iteration += 1
return coefficients

def calculate_error(predicted_data):
    # Обчислюємо середньоквадратичну помилку
    print(predicted_data)
    mse = mean_squared_error(predicted_data.iloc[:,3],
predicted_data['прогнозоване значення'])
    print(f'Mean Squared Error: {mse}')
    # Будуємо графік
    plt.figure(figsize=(10, 6))
    plt.plot(predicted_data.index, predicted_data.iloc[:,3], label='Тестові значення')
    plt.plot(predicted_data.index, predicted_data['прогнозоване значення'],
label='Прогнозовані значення')
    plt.xlabel('Індекси')
    plt.ylabel('Значення')
    plt.title('Прогнозовані та тестові значення')
    plt.legend()

```

```

plt.annotate(f'MSE: {mse}', xy=(0.05, 0.95), xycoords='axes fraction',
fontsize=12, bbox=dict(facecolor='white', alpha=0.5))
plt.savefig("mse.jpg")
# plt.show()
def main(train_path: str, test_path: str) -> None:
    # Обробка навчальних даних
    train_file_path = train_path
    train_data = proceed_train_data(train_file_path)
    train_data = clean_data(train_data)
    # Обробка тестових даних
    test_file_path = test_path
    test_data = pd.read_csv(test_file_path, sep=";")
    test_data = process_test_data(test_data)
    test_data = clean_data(test_data)
    # Ініціалізація коефіцієнтів
    num_coefficients = len(train_data.columns) - 3 # Вираховуємо кількість
коефіцієнтів
    coefficients = initialize_coefficients(num_coefficients)
    # Навчання моделі
    learning_rate = 0.01
    regularization_lambda = 0.1
    max_iterations = 1000
    tolerance = 1e-5
    coefficients = train_model(train_data, coefficients, learning_rate,
regularization_lambda, max_iterations, tolerance, train_data)
    # Прогнозування значень для тестових даних
    predicted_data = predict_values(test_data, coefficients)
    # Обчислення помилки та вивід результатів
    calculate_error(predicted_data)

```

## ДОДАТОК Б

Файл бібліотек до програми під назвою **req.txt**

contourpy==1.2.1

cycler==0.12.1

et-xmlfile==1.1.0

fonttools==4.51.0

joblib==1.4.2

kiwisolver==1.4.5

matplotlib==3.8.4

natsort==8.4.0

numpy==1.26.4

openpyxl==3.1.2

packaging==24.0

pandas==2.2.2

pillow==10.3.0

pyarsing==3.1.2

python-dateutil==2.9.0.post0

pytz==2024.1

scikit-learn==1.4.2

scipy==1.13.0

six==1.16.0

threadpoolctl==3.5.0

tzdata==2024.1