

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій


Комплексна магістерська кваліфікаційна робота на тему

**АНАЛІЗ ТА ПЕРЕДБАЧЕННЯ ДОБОВОГО ГЕНЕРУВАННЯ
ЕЛЕКТРОЕНЕРГІЇ ФОТОЕЛЕКТРОСТАНЦІЄЮ. ЧАСТИНА 3. АНАЛІЗ
ВПЛИВУ СОНЯЧНОЇ АКТИВНОСТІ**


Виконав: студент 2 курсу, групи 2ІСТ-22м
спеціальності 126 – «Інформаційні системи
та технології»

 Ігор КРУЧАЙ

Керівник: д.т.н., проф. каф. САІТ

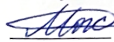
 Олександр МОКІН
« 08 » 12 2023 р.

Опонент: д.т.н., проф. каф. КН

 Ярослав ІВАНЧУК
« 12 » 12 2023 р.

Допущено до захисту

Завідувач кафедри САІТ

 д.т.н., проф. Віталій МОКІН


« 08 » 12 2023 р.

Вінниця ВНТУ – 2023 рік

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій
Рівень вищої освіти – II-й (магістерський)
Галузь знань – 12 Інформаційні технології
Спеціальність – 126 Інформаційні системи та технології
Освітньо-професійна програма – Інформаційні технології аналізу даних та зображень

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

 д.т.н., проф. Віталій МОКІН

« 01 » 09 2023 р.



ЗАВДАННЯ НА КОМПЛЕКСНУ МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Кручаю Ігорю Степановичу

- Тема роботи: “Аналіз та передбачення добового генерування електроенергії фотоелектростанцією. Частина 3. Аналіз впливу сонячної активності”, керівник роботи: Олександр МОКІН., д.т.н., проф. каф. САІТ, затверджені наказом закладу вищої освіти від « 14 » 09 2023 року № 047
- Строк подання студентом роботи « 30 » 11 2023 року
- Вихідні дані до роботи:
 - Kaggle Dataset „EnergyAndSolarRadiation”
<https://www.kaggle.com/datasets/igorkruchay/energyandsolarradiation>, наданий з дозволу адміністрації фотоелектричної станції, розташованої у Вінницькій області.
- Зміст текстової частини:
 - загальна характеристика об’єкту досліджень;
 - вибір оптимальних налаштувань для розв’язання поставленої задачі;
 - системний аналіз та передбачення добового генерування електроенергії фотоелектростанцією;
 - економічна частина.
- Перелік ілюстративного матеріалу:
 - добові графіки генерації електроенергії;
 - добові графіки фіксації сонячної активності;
 - графіки кореляційних матриць між ознаками;
 - блок-схема алгоритму інформаційної технології;
 - вибір оптимальної моделі передбачення;
 - графіки передбачень моделей для різних наборів даних.







– UML-діаграма Use Case використання інформаційної технології.

6. Консультанти розділів МКР

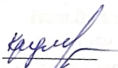
Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
4	Наталія БУРСНІКОВА, д.е.н., проф. каф. ЕПВМ	29.10 	05.11 

7. Дата видачі завдання « 04 » 09 2023 року

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва та зміст етапу	Термін виконання		Примітка
		початок	закінчення	
1	Аналіз предметної області	01.09.23	30.09.23	
2	Вибір оптимальних інформаційних технологій	01.10.23	09.10.23	
3	Розроблення інформаційної технології	10.10.23	24.10.23	
4	Тренування та тестування моделей машинного навчання	25.10.23	28.10.23	
5	Економічна частина	29.10.23	03.11.23	
6	Оформлення матеріалів до захисту МКР	04.11.23	25.11.23	

Студент



Ігор КРУЧАЙ

Керівник роботи



Олександр МОКІН

АНОТАЦІЯ

УДК 004.08

Кручай І.С. Аналіз та передбачення добового генерування електроенергії фотоелектростанцією. Частина 3. Аналіз впливу сонячної активності. Комплексна магістерська кваліфікаційна робота зі спеціальності 126 – інформаційні системи та технології, освітньо-професійна програма – інформаційні технології аналізу даних та зображень. Вінниця: ВНТУ, 2023. 108 с.

На укр. мові. Бібліогр.: 42 назв; рис.: 54; табл.: 8.

Магістерська кваліфікаційна робота присвячена розробці інформаційної технології аналізу та передбачення добового генерування електроенергії фотоелектростанцією з урахуванням впливу сонячної активності. В якості вхідних даних використано датасет з об'ємами генерації та показниками сонячної активності для фотоелектричної станції, що розташована на території Вінницької області. В роботі проведений аналіз предметної області, визначені існуючі проблеми та актуальність їх вирішення. Розглянуті аналоги та обґрунтована необхідність нового дослідження. Проведено розвідувальний аналіз вхідних даних, вибрані оптимальні інформаційні технології машинного навчання для побудови моделей передбачення. Побудовано декілька моделей та обрано найкращу з них. Зроблено висновки щодо її ефективності.

Ілюстративна частина складається з 12 плакатів із результатами моделювання.

У розділі економічної частини розглянуто питання про доцільність розробки та впровадження інформаційної технології на реально працюючому об'єкті.

Ключові слова: фотоелектрична станція, сонячна активність, датасет, модель передбачення.

ABSTRACT

Kruchai I.S. Analysis and prediction of daily electricity generation by a solar power station (complex MCR). Part 3. Analysis of the impact of solar activity. Comprehensive master's qualification thesis on specialty 126 - information systems and technologies, educational and professional program - information technologies of data and image analysis. Vinnytsia: VNTU, 2023. 108 p.

In Ukrainian language. Bibliogr .: 42 titles; fig .: 54; table: 8.

The master's qualification work is dedicated to the development of an information technology for the analysis and prediction of daily electricity generation by a photovoltaic power station, taking into account the influence of solar activity. A dataset with generation volumes and solar activity indicators for a photovoltaic station located in the Vinnytsia region is used as input data. The paper includes an analysis of the subject area, identification of existing problems, and the relevance of their solutions. Analogues are considered, and the need for new research is justified. A reconnaissance analysis of input data is conducted, optimal machine learning technologies for prediction model construction are selected, and several models are built, with the best one chosen. Conclusions are drawn regarding its effectiveness.

The illustrative part consists of 12 posters with the modeling results.

In the economic section, the feasibility of developing and implementing the information technology on a real working object is discussed.

Key words: Photovoltaic station, solar activity, dataset, prediction model.

ЗМІСТ

ВСТУП.....	4
1 ЗАГАЛЬНА ХАРАКТЕРИСТИКА ОБ’ЄКТУ ДОСЛІДЖЕНЬ.....	8
1.1 Аналіз предметної області, існуючі проблеми та актуальність їх вирішення	8
1.2 Огляд і опис методів прогнозування та вибір оптимальних інформаційних технологій.....	17
1.3 Огляд та аналіз аналогів	26
1.4 Висновки	28
2 ВИБІР ОПТИМАЛЬНИХ НАЛАШТУВАНЬ ДЛЯ РОЗВ’ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ	29
2.1 Веб платформа Kaggle.....	29
2.2 Мова програмування Python.....	31
2.3 Дослідження та аналіз вхідних даних.....	34
2.4 Розвідувальний аналіз даних та вибір оптимальних налаштувань.....	49
2.5 Висновки	57
3 СИСТЕМНИЙ АНАЛІЗ ТА ПЕРЕДБАЧЕННЯ ДОБОВОГО ГЕНЕРУВАННЯ ЕЛЕКТРОЕНЕРГІЇ ФОТОЕЛЕКТРОСТАНЦІЄЮ	59
3.1 Розділення вхідних даних на тренувальні, тестові та валідаційні.....	59
3.2 Побудова моделей передбачення генерування електроенергії та вибір найкращої з них	62
3.3 Висновки	74
4 ЕКОНОМІЧНА ЧАСТИНА	76
4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки.....	76
4.2 Розрахунок витрат на проведення науково-технічної роботи	81
4.2.1 Витрати на оплату праці.....	81
4.2.2 Відрахування на соціальні заходи.....	84

4.2.3	Сировина та матеріали	84
4.2.4	Розрахунок витрат на комплектуючі.....	85
4.2.5	Спецустаткування для наукових (експериментальних) робіт	85
4.2.6	Програмне забезпечення для наукових (експериментальних) робіт ...	85
4.2.7	Амортизація обладнання, програмних засобів та приміщень	86
4.2.8	Паливо та енергія для науково-виробничих цілей	87
4.2.9	Службові відрядження.....	88
4.2.10	Витрати на роботи, які виконують сторонні підприємства, установи і організації	88
4.2.11	Інші витрати	89
4.2.12	Накладні (загальновиробничі) витрати.....	89
4.3	Розрахунок економічної ефективності науково-технічної розробки за її можливої комерціалізації потенційним інвестором.....	90
4.4	Висновки	97
ВИСНОВКИ		98
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....		100
Додаток А (обов'язковий). Технічне завдання		105
Додаток Б (обов'язковий). Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень		107
Додаток В (довідковий). Лістинг програми		108
Додаток Г (обов'язковий). Ілюстративна частина.....		112

ВСТУП

Актуальність теми. Практично для всіх країн світу є дуже актуальним питання, що стосується використання нетрадиційних відновлюваних джерел енергії. Для тих країн, що відносяться до розвинутих, але імпортують паливно-енергетичні ресурси, це безумовно стосується забезпечення їх енергетичної безпеки. Для тих країн, що мають свої власні запаси палива, це стосується поліпшення екологічної ситуації в країні. Для тих країн, які розвиваються, але не є достатньо розвинутими в економічному сенсі, це один із способів покращити економічні показники, а також соціально-побутові умови населення цієї країни. Нетрадиційні відновлювані джерела енергії, як правило, розглядають як істотне доповнення до традиційних.

Сьогодні сонячна енергетика являє собою один із сучасних, прогресуючих та перспективних напрямів використання енергії відновлюваних джерел. Перевагою сонячної енергетики можна вважати безпосереднє перетворення енергії сонячної випромінення (радіації) в енергію електричну. Потенційні можливості сонячної енергетики надзвичайно великі. Якщо використати всього лише 0,0125% всієї кількості енергії Сонця, це могло б забезпечити всі сьогоденні потреби світової енергетики. Використання 0,5% енергії Сонця – повністю покрило би потреби людства на перспективу. Однак існують і перешкоди. Найбільш суттєві, це нерівномірність сонячного опромінення, не завжди висока інтенсивність падаючих променів (залежність від року, сезону), а також іноді значна вартість пристроїв, що можуть перетворювати енергію сонця в електричну енергію [1].

Україна має площу близько 603,000 км² та розташована в різних кліматичних зонах. Потенціал сонячної енергії в Україні є достатньо високим для того, щоб впроваджувати фото-електроенергетичне обладнання практично на всій її території. Якщо поррахувати технічно-досяжний енергетичний потенціал сонячної енергії в Україні за один рік, то він є еквівалентним від 6 до 6,5 млн т. умовного палива. Його використання дозволило би заощадити

біля 5-6 млрд. м³ природного газу. Якщо збільшити споживання сонячної енергії, то в багатьох місцях, це розв'язало би проблеми гарячого водопостачання в теплу пору року [2].

В Україні енергоринок відновлюваних джерел енергії стимулюється пільговими тарифами. Останнім часом в Україні прискорився процес встановлення та експлуатації об'єктів генерації з відновлюваних джерел енергії. Це в першу чергу зумовлено схемою «зелених тарифів», яка є досить приваблива, а також зниженню вартості новітніх технологій, покращення бізнес-клімату в середині країни, вагомого економічного та технічного потенціалу країни в цілому. Якщо взяти період з третього кварталу 2016 року до кінця другого кварталу 2018 року, потужність сонячної генерації (фотовольтаїка) – зросла на близько 78 - 79 % [3].

Популяризація відновлюваних джерел енергії призвела до стрімкого підвищення їх частки в енергосистемі України. З іншого боку, така інтенсивність розвитку та впровадження мережевих сонячних електростанцій, призводить до значних проблем. В першу чергу сюди відносяться проблеми, що стосуються управління та надійності роботи сонячних електричних систем внаслідок випадкового характеру природних та погодних явищ, таких як рівень сонячної активності, хмарність, інтенсивність опадів, туман та інші. Мінливість погоди, призводить до складності передбачення кількості енергії, яку може генерувати фото електростанції, що в свою чергу додає складності в балансуванні електричної мережі країни в загалом та може призводити до накладання штрафних санкцій до електрогенеруючих компаній [2].

Саме тому прогнозування та передбачення генерації фото електростанціями є досить важливим питанням для енергетиків. Маючи ефективну модель передбачення або прогнозування, можна підвищити надійність роботи електричних систем країни, покращити балансування електричної системи, робити більш ефективно планування регламентних робіт

та обслуговування обладнання, а також підвищити економічну ефективність фото електростанцій за рахунок зменшення штрафних санкцій.

Метою роботи є підвищення точності передбачення генерації електроенергії фотоелектричною станцією за допомогою методів машинного навчання та системного аналізу, з урахуванням сонячної активності.

Для досягнення мети необхідно розв'язати наступні задачі:

- на основі отриманого датасету дослідити зміни генерації електроенергії в залежності від показників сонячної активності;
- вибрати оптимальні інформаційні технології машинного навчання для побудови моделей передбачення;
- побудувати декілька моделей передбачення, використовуючи різні методи;
- провести навчання моделей на тренувальному наборі даних;
- перевірити точність моделей на валідаційному наборі даних;
- вибрати модель з найменшою похибкою;
- за допомогою обраної моделі здійснити передбачення генерації електроенергії на тестових даних та оцінити ефективність.

Об'єктом дослідження є процес аналізу даних та передбачення генерації електроенергії фотоелектричною станцією з урахуванням сонячної активності.

Предметом дослідження є методи машинного навчання, які ґрунтуються на сучасних інформаційних технологіях та які використовуються для побудови моделей передбачення генерації електроенергії фотоелектричною станцією з урахуванням сонячної активності.

Наукова новизна. Дістала подальший розвиток інформаційна технологія передбачення добового генерування електроенергії фотоелектричною станцією, за рахунок використання методів і моделей машинного навчання та оригінальних ознак на основі сонячної активності, що дозволяє підвищити точність цього передбачення.

Практичне значення. Результат магістерської кваліфікаційної роботи має допомогти виробникам фото електричної енергії покращити балансування, підвищити надійність роботи їх генеруючих систем, а також мати змогу робити більш ефективно планування регламентних робіт та обслуговування електричного обладнання.

Апробація результатів магістерської кваліфікаційної роботи. Результати магістерської роботи доповідались на ЛІІ Науково-технічній конференції факультету інтелектуальних інформаційних технологій та автоматизації ВНТУ (м. Вінниця, 2023 р.).

Публікації результатів магістерської кваліфікаційної роботи. Опубліковано тези на ЛІІ Науково-технічній конференції факультету інтелектуальних інформаційних технологій та автоматизації ВНТУ (м. Вінниця, 2023 р.) [4]

1 ЗАГАЛЬНА ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕНЬ

1.1 Аналіз предметної області, існуючі проблеми та актуальність їх вирішення

Процес виробництва електроенергії на сонячній станції є повністю екологічним і не створює забруднюючих елементів для навколишнього середовища, а також є одним з найефективніших відновлюваних джерел енергії, які існують на даний момент.

Завдяки перевагам сонячної енергії в порівнянні з енергією, виробленою з вичерпного палива або невідновлюваних джерел, сонячні електростанції є ключовим інструментом для розробки нової моделі довгострокового сталого виробництва, яка є повністю екологічною. Існує два основних типи сонячних електростанцій: сонячні фотоелектричні станції та сонячні теплові електростанції. Фотоелектричні станції є універсальними, а отже найбільш поширеними в світі [3]. На рисунку 1.1 показані сонячні панелі «Слобода-Бушанської» фотоелектричної станції, що розташована у Вінницькій області.



Рисунок 1.1 – «Слобода-Бушанська» фотоелектрична електростанція у Вінницькій області

Робота сонячної фотоелектричної установки заснована на фотонах і світловій енергії сонячних променів. Типи сонячних панелей, які використовуються в цих типах об'єктів, також відрізняються. У той час як сонячні теплові електростанції використовують колектори, фотоелектричні електростанції використовують панелі, що складаються з фотоелектричних сонячних елементів, виготовлених із кремнію (монокристалічні або полікристалічні сонячні панелі) або інших матеріалів з фотоелектричними властивостями (аморфні сонячні панелі).

Завдяки матеріалам, з яких вони виготовлені, коли сонячні промені потрапляють на них безпосередньо, електрони, що містяться у більш зовнішніх шарах фотоелектричних елементів, поглинають випромінювання та генерують постійний струм. Щоб побудувати сонячну фотоелектричну станцію, ці сонячні батареї з'єднуються послідовно, щоб бути інтегрованими в єдиний модуль, утворюючи фотоелектричні панелі.

У свою чергу, сонячні панелі з'єднані паралельно, утворюючи струни, з'єднані з інвертором струму, де постійний струм, що надходить від фотоелементів, перетворюється на змінну енергію. Потім електроенергія направляється до трансформатора, в якому її напруга та інтенсивність адаптовані так, щоб її можна було транспортувати через лінії електричної мережі до центрів споживання. Типова схема фотоелектричної станції наведена на рисунку 1.2 [5].



Рисунок 1.2 – Типова схема фотоелектричної станції

Що стосується Сонця, то на поверхню земної кулі попадає сонячне випромінювання. Енергія цього випромінювання становить близько $81000 \cdot 10^6$ МВт. Більша його частина припадає на водну поверхню. Менша частина, $27000 \cdot 10^6$ МВт, на континенти. Кількість енергії також коливається в залежності від сезону. У період з початку квітня до кінця вересня надходить близько 80% цієї енергії. Всесвітнє використання всіх різновидів енергії становить близько $10 \cdot 10^6$ МВт в рік. Якщо провести розрахунки, побачимо, що за одну годину від сонця наша планета одержує майже стільки ж енергії, скільки людству необхідно для використання впродовж року. Розподіл сонячного випромінювання не є однаковим і залежить від географічного розташування, що видно на рисунку 1.3 [6].

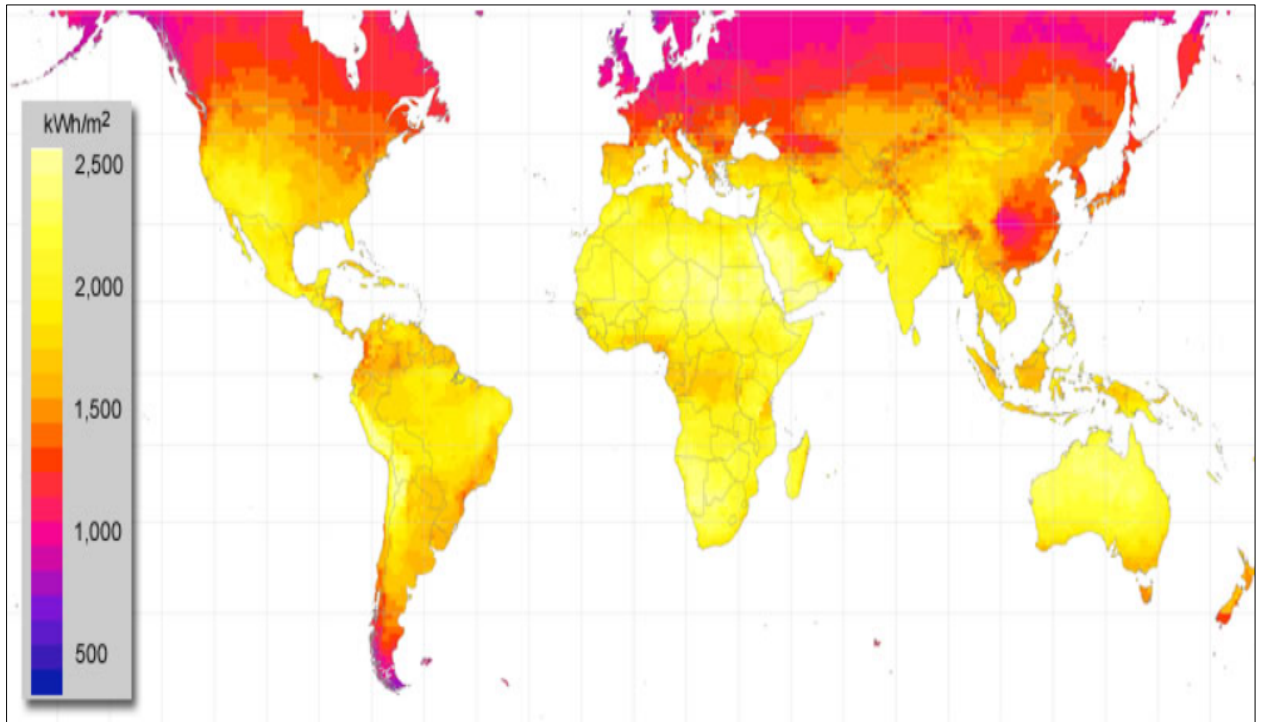


Рисунок 1.3 – Різниця розподілення сонячного випромінювання в залежності від географічного положення

Виникає логічне питання, чому при такому величезному потенціалі дармової енергії сонця, її настільки важко перетворити на енергію, яка доступна для використання. Та, чому отримана від сонця енергія є такою дорогою в порівнянні з енергією отриманою від традиційних джерел. Відповідями на ці питання можна вважати малу щільність сонячної енергії; нерівномірність та невисоку інтенсивність падаючих променів; значна вартість пристроїв перетворення сонячної енергії.

Якщо говорити про Україну, то її потенціал для розвитку та використання сонячної енергії є достатньо високим, щоб впроваджувати системи фото-електроенергетичного обладнання практично на всій її території. Сонячне випромінювання за рік в Україні складає приблизно 3500–5200 МДж/м². Якщо дивитись окремо по регіонах, картина буде різною. У північних регіонах сезонний період для активного використання сонячної енергії може продовжуватись з квітня по вересень, тоді як у південних регіонах - з початку березня по кінець жовтня, що становить близько 1900–

2400 годин в рік. Якщо взяти загальне середньорічне сонячне випромінювання, то воно змінюється від 1070 кВт·год/км² на місяць в північних регіонах до 1400 кВт·год/км² в південних регіонах країни [7].

На рисунку 1.4 показано розподілення сонячного випромінювання в Україні.

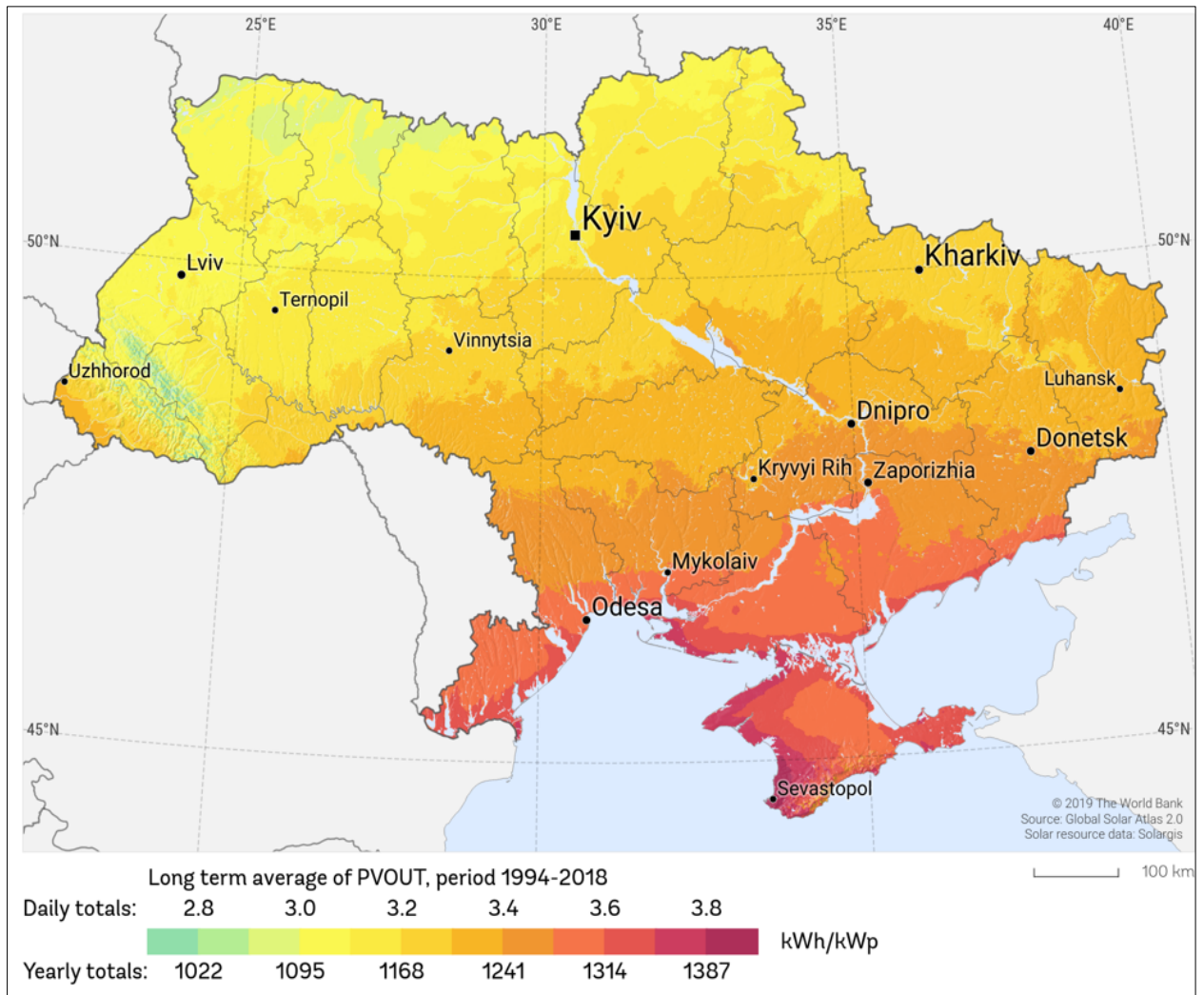


Рисунок 1.4 – Розподілення сонячного випромінювання в Україні

У цілому територія України відноситься до місць з середньою інтенсивністю сонячного випромінювання. Якщо узагальнити показники, то на території України, за рівнем інтенсивності сонячного випромінювання можна виділити чотири окремі зони, які показані на рисунку 1.5. У першій та другій

зонах, де інтенсивність найбільша, знаходяться усі південні області України. В третій зоні, з середньою інтенсивністю, знаходяться західні області. В четвертій зоні знаходиться приблизно половина території нашої країни. Найменш придатна для використання сонячної енергії четверта зона.

В першій зоні найбільша величина надходження сонячного випромінення – 1350 - 1400 кВт·год/км² в рік, в четвертій – найменша, приблизно 1000 кВт·год/км² в рік. Для другої зони - 1250 кВт·год/км². Для третьої - 1150 кВт·год/км² за рік [5].

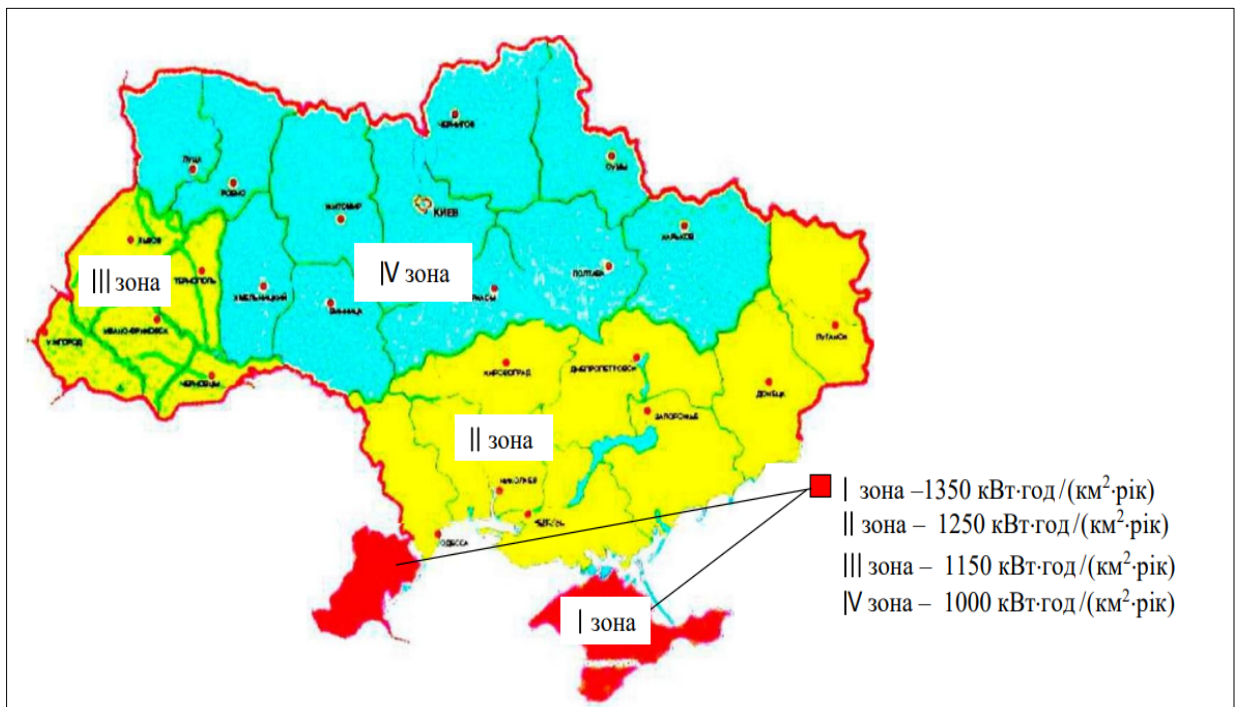


Рисунок 1.5 – Поділ території України на зони, згідно інтенсивності сонячного випромінення

Сонячна активність являє собою електромагнітні хвилі різної довжини. Діапазон хвиль, що досягають поверхні Землі, досить широкий: від радіохвиль до рентгенівських променів. Сонячна активність має циклічну природу і змінюється із року в рік. Це має назву – сонячні цикли (рис 1.6). Є роки з доволі високою сонячною активністю та навпаки з доволі низькою. В 1960 році, вчені

фіксували найбільшу сонячну активність з початку її дослідження. Вона була пов'язана з активними та хаотичними спалахами на Сонці.

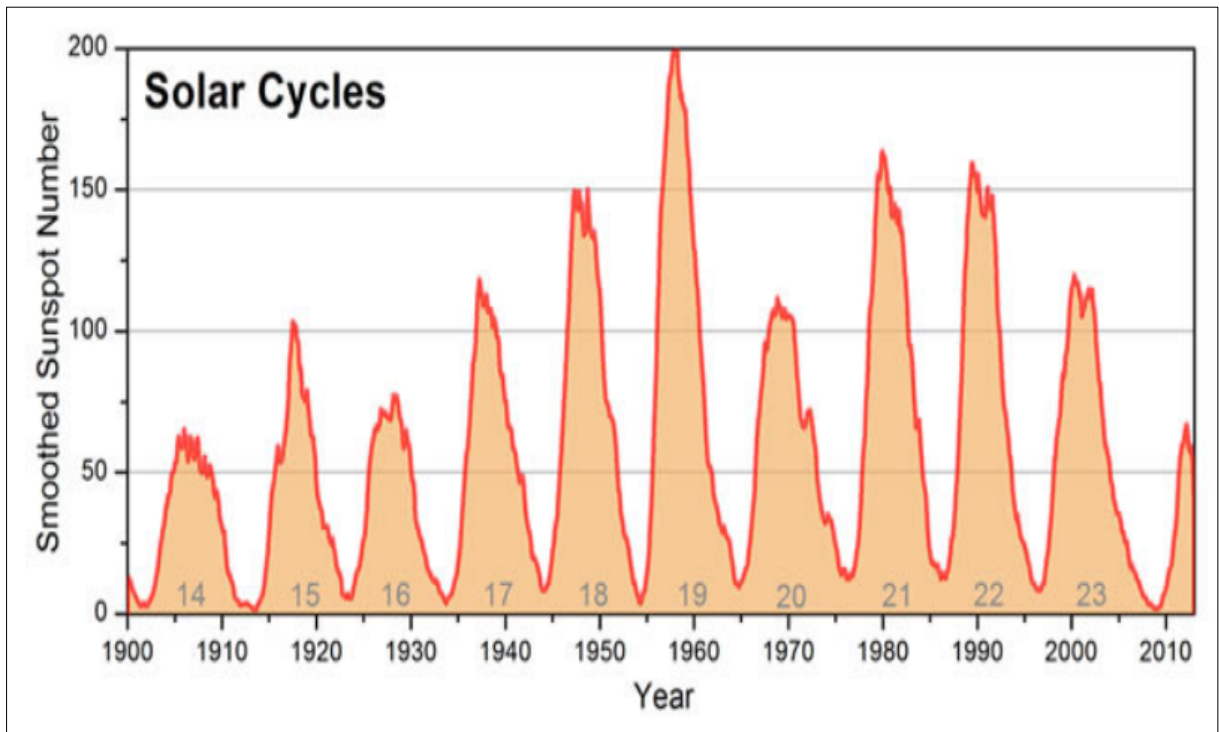


Рисунок 1.6 – Сонячні цикли

Радіація, що досягає поверхні нашої планети, може бути прямою, дифузною чи відбитою (рис. 1.7):

- Пряма – це те опромінення яке безпосереднього попадає на поверхню;
- Дифузна – пов'язана з великою кількістю природних перешкод на шляху сонячного опромінення, що приводить до його розсіювання: озоновий шар атмосфери, хмари, водяна пара та пил у повітрі;
- Відбита – пов'язана з тою частиною опромінення, що відбивається від поверхні землі.

Іноді дифузну та відбиту радіацію не відокремлюють одну від іншої та називають просто - розсіяною.

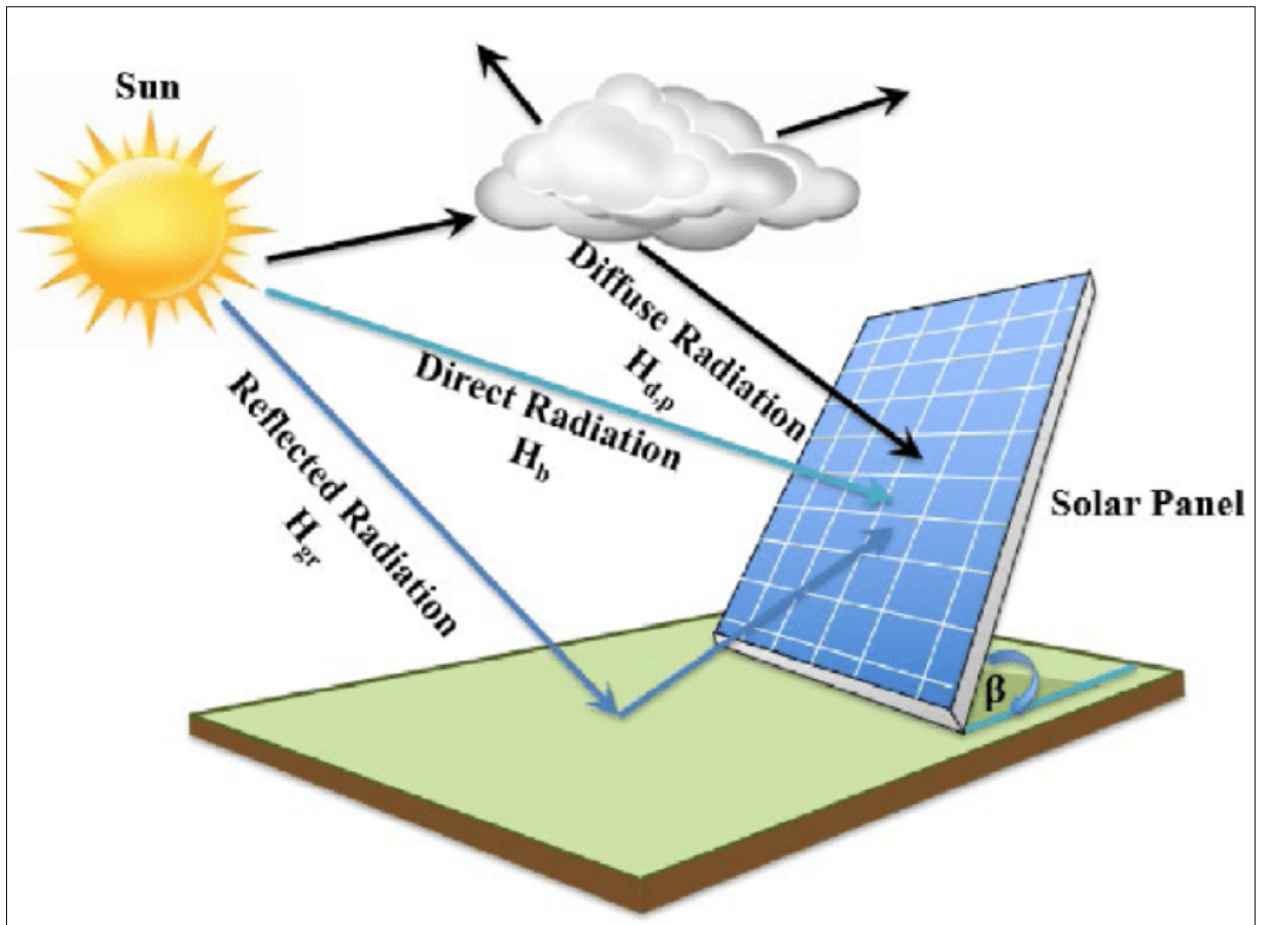


Рисунок 1.7 – Пряма, дифузійна та відбита сонячна радіація

В залежності від сонячної активності, змінюється величина густини прямої, дифузної та відбитої сонячної радіації, рисунок 1.8. Є також залежність від прозорості атмосфери, характеристик земної поверхні, широти місцевості, а також від пори року та часу доби.

Саме тому, кількість сонячної радіації що за рік потрапляє на 1 м^2 земної поверхні, значно відрізняється для різних регіонів України. Очевидним є збільшення кількості сонячних днів, кількості сонячної енергії, а відповідно і густини сонячної радіації на 1 м^2 поверхні землі в напрямку з Півночі на Південь [6].

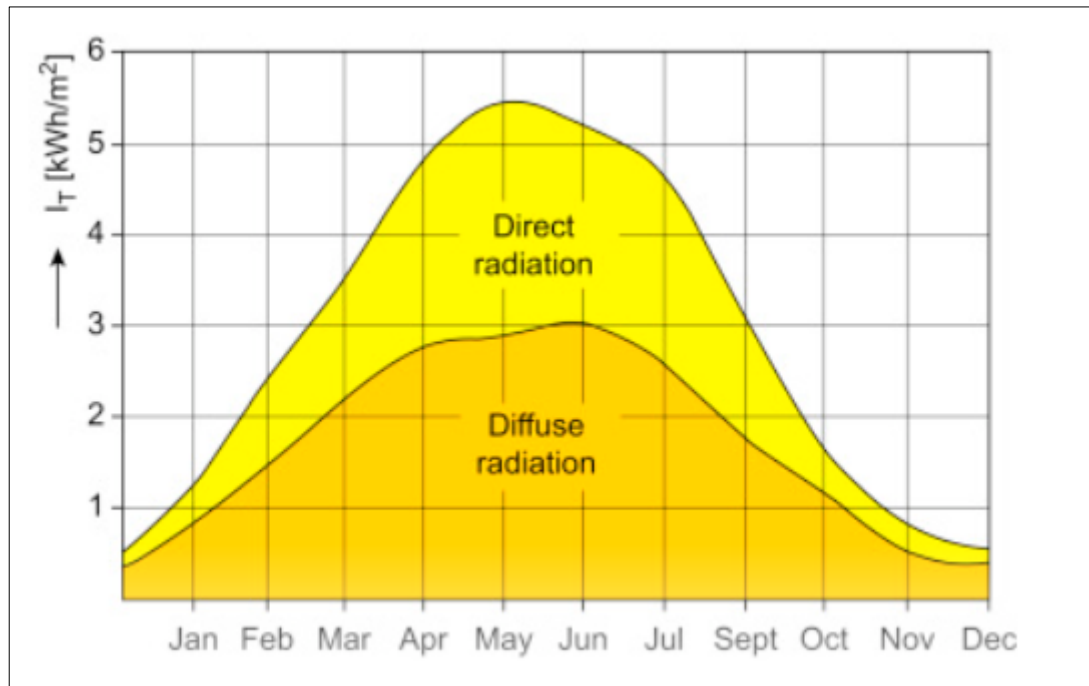


Рисунок 1.8 – Співставлення показників прямої та дифузійна сонячної радіації

Отже, оскільки є багато факторів, що впливають на сонячну радіацію, відповідно генерація електроенергії, що відбувається завдяки сонячним панелям, є нерівномірною. Велика частка нерегульованих або погано регульованих нестабільних джерел енергії в енергосистемі може призвести до її значного розбалансування. Це є проблемою, тому що в енергосистемі дуже важливо забезпечити балансову стійкість. До того ж, як відомо, в Україні електрогенеруючі компанії обкладаються досить великими штрафами за недотримання балансу з вироблення електроенергії [8]. В зв'язку з цим, на сьогоднішній день проблеми прогнозування або передбачення генерації фотоелектричними станціями є актуальним та важливими.

1.2 Огляд і опис методів прогнозування та вибір оптимальних інформаційних технологій

Прогнозування - це процес формування оцінок щодо майбутніх подій чи умов на основі минулих та поточних даних. Прогнозування використовується в широкому спектрі галузей, включаючи бізнес, економіку, фінанси, метеорологію та інженерію. Мета прогнозування полягає в тому, щоб приймати обґрунтовані рішення щодо майбутнього на основі минулих тенденцій та поточних даних. Прогнозування використовує статистичні моделі та інші кількісні методи для аналізу історичних даних та вироблення прогнозів щодо майбутніх тенденцій.

Передбачення - це оцінювання майбутніх результатів чи подій на основі неповних або невизначених даних. Передбачення використовується в різних галузях, таких як інженерія, психологія, соціальні науки та медицина, щоб приймати обґрунтовані рішення щодо індивідуальної поведінки чи результатів. Метою передбачення є формування висновків про майбутні події на основі обмеженої інформації, часто застосовуючи суб'єктивні чи якісні методи.

Основна різниця між прогнозуванням та передбаченням полягає в рівні невизначеності, що включена у процес. Прогнозування ґрунтується на історичних даних та статистичних моделях для створення прогнозів щодо майбутніх тенденцій, тоді як передбачення використовує неповну чи невизначену інформацію для висновків про майбутні події. Прогнозування зазвичай використовується для прийняття рішень щодо подій або тенденцій великого масштабу, тоді як передбачення використовується для прийняття рішень щодо індивідуальної поведінки чи результатів [9].

Від точності прогнозування або передбачення напряму залежить фінансове благополуччя електрогенеруючої компанії, оскільки суттєве відхилення графіка генерування від заявленого може призвести до значних

збитків. Саме тому грамотне і точне прогнозування відіграє важливу роль в успішному впровадженні систем “зеленої” енергетики в енергетичну галузь нашої країни [10].

Для прогнозування процесів генерації електричної енергії на сьогодні використовують багато різних підходів та методів. Якщо проаналізувати методики розробки математичних моделей, алгоритмів для їх реалізації, програмних засобів, що використовуються для розв'язку задач короткострокового (доба, декілька діб, тиждень), середньострокового (від будь-якої дати до кінця поточного місяця або місяця в цілому) та довгострокового (на майбутні місяці, квартал, рік тощо) прогнозування електроспоживання і активного навантаження в електричній мережі показує, що на вибір відповідного методу прогнозування можуть впливати наступні чинники: необхідна точність, інтервал прогнозування, доступність та якість даних, поведінка прогнозованого процесу [11].

Основні вимоги, що ставляться до сучасних комп'ютерних моделей передбачення та прогнозування:

- адаптивність – являє собою можливість коригування моделі для досягнення необхідної точності передбачення/прогнозування;
- стійкість – показує наскільки математична прогностична модель є стійкою до помилок у вхідних даних або до часткового відсутності деякої їх частини;
- інтервальність – дає можливість для обчислення величини передбачення або прогнозу для вибраного часового інтервалу (від будь-якої довільної дати до кінця розрахункового періоду);
- робастність – показує наскільки надійним є результат прогнозу; як правило досягається за рахунок застосування оптимальних моделей, та стійкість до можливих помилок у вхідних даних;
- інтерактивність – можливість введення коригувальної інформації в модель в інтерактивному режимі кінцевим користувачем;

- повнота – показує можливість реалізації деяких додаткових функцій, що властиві об'єкту прогнозування для досягнення необхідної точності передбачення/прогнозування;
- рекурсивність – являє собою можливість додаткового навчання моделі впродовж надходження нових даних для підвищення якості передбачення/прогнозування;
- економність – мінімізація витрат на використання засобів автоматизації та інформаційних ресурсів.

Для вирішення конкретної практичної задачі, вибір оптимального методу чи методики, це досить непроста задача, як може здаватись на перший погляд. При створенні моделей прогнозування, в багатьох енергетичних системах, використовуються статистичні методи аналізу такі як часові ряди. Часові ряди (динамічні ряди) являють собою впорядковані послідовності даних спостережень за процесом (вимірів), що змінюються в часі [12].

Відомо достатню кількість математичних стохастичних методів прогнозування графіків електричного навантаження, які базуються на аналізі часових рядів регресійних моделей. При їх практичній реалізації, для зменшення похибки розрахунків потрібно збирати, обробляти та зберігати значні обсяги первинних даних. До яких відноситься фактичного виробництво електричної енергії. Крім того, додатково необхідно зберігати великий обсяг даних, що призначені для визначення та подальшого виправлення помилок. Для забезпечення коректності даних, врахування сторонніх факторів, що можуть впливати на кінцевий результат прогнозу [13]. В контексті даної роботи, до таких факторів можна віднести сонячну активність, хмарність, вологість та інші природні чинники.

Дослідження та аналіз інформації щодо різних методів прогнозування, показує, що на сьогодні немає єдиного універсального методу чи методики, яка дозволила би вирішити проблеми прогнозування випадкових процесів різної природи. Однак, вироблені підходи, для вирішення конкретних

прикладних задач, дають можливість побудови моделі прогнозування, що забезпечує прийнятну для практики достовірність і точність. Серед різноманіття методів побудови моделей прогнозування та передбачення з використанням методів машинного навчання, можна виділити наступні [14]:

- нейронні мережі;
- Баєсові мережі;
- дерева рішень.

В нейронних мережах всі обчислення будуються за допомогою взаємозв'язаних груп штучних нейронів, які можуть обробляти інформацію використовуючи конективістський підхід до обчислень. Сучасні нейронні мережі є нелінійними статистичними інструментами для моделювання даних. Зазвичай їх застосовують для пошуку закономірностей в даних, моделювання нечітких та складних взаємозв'язків між входами та виходами або для виявлення в невідомому спільному розподілі ймовірностей спостережуваних величин деякої статистичної структури. Нейронні мережі можуть бути апаратними та програмними. В апаратних мережах нейрони подано фізичними складовими. В програмних мережах роль нейронів відіграють комп'ютерні моделі, а також можуть використовуватись різноманітні алгоритми навчання, запроваджуватись різні топології. Існує велика кількість різновидів нейронних мереж. Кожен різновид використовується в різних сферах та призначений для вирішення певних задач. Деякі з основних типів нейронних мереж:

- Перцептрон (Perceptron) є основним блоком для багатьох інших нейронних мереж. Використовується для бінарної класифікації;
- Багатошаровий перцептрон (Multilayer Perceptron, MLP) складається з декількох шарів перцептронів і використовується для різноманітних завдань, включаючи класифікацію та регресію;
- Зворотнє поширення (Backpropagation) нейронна мережа використовується для навчання багатьох інших типів нейронних мереж,

зокрема MLP;

- Рекурентні нейронні мережі (Recurrent Neural Networks, RNN) призначені для обробки даних, що є послідовними, таких як часові ряди або текст;
- Згорткові нейронні мережі (Convolutional Neural Networks, CNN) ефективні для обробки зображень і використовують згорткові шари для виявлення патернів у зображеннях;
- Довга короткострокова пам'ять (Long Short-Term Memory, LSTM), є видом рекурентних нейронних мереж, які здатні зберігати та використовувати інформацію на тривалий термін;
- Мережі довгої пам'яті і короткочасної пам'яті (Long Short-Term Memory, LSTM) використовуються для роботи з послідовними даними і мають покращену здатність до вирішення проблем втрати градієнту;
- Автокодері (Autoencoders), використовуються для витягнення важливих ознак з даних та стиснення і розгортання інформації;
- Мережі денної пам'яті (Memory Networks), використовуються для вирішення завдань, пов'язаних з розумінням та запитам до пам'яті;
- Мережі асоціативної пам'яті (Hopfield Networks, Boltzmann Machines), використовуються для розв'язання завдань асоціативної пам'яті і оптимізації;
- Трансформери (Transformers), використовуються для обробки послідовних даних та роботи з прикладами з різних контекстів;
- Генеративні адверсаріальні мережі (Generative Adversarial Networks, GAN), використовуються для генерації нових даних, таких як зображення або тексти;
- Трансформери для обробки мовлення (BERT, GPT, T5 і т. д.) використовуються для завдань обробки природної мови, таких як машинний переклад, сентимент-аналіз та інші [14-15].

Баєсові мережі - ймовірнісні графові моделі, що являють собою набір

випадкових величин та їх умовних незалежностей один від одного через ациклічний спрямований граф. Мережі Баєса все ширше знаходять застосування в інформаційних системах обробки статистичних даних, що представлені часовими зрізами або часовими рядами. Також можуть застосовуватись в системах обробки де обробляються якісні дані, що представлені лінгвістичними змінними, інтервальними значеннями, експертними оцінками. Більше того, ці типи мереж також застосуються в системах автоматичного розпізнавання мовних сигналів, класифікації даних різної природи, маркетингу, бізнесі та інших сферах діяльності. Загалом Байєсові мережі дають можливість встановлювати причинно-наслідкові зв'язки між подіями та визначати ступінь імовірності настання тієї чи іншої ситуації при отриманні інформації, що стосується зміни стану будь-якого із вузлів мережі. Наскільки успішним буде застосування даного методу моделювання та наскільки правильним буде формування статистичного висновку, безумовно залежить від багатьох чинників. А саме: вміння чітко формулювати задачу, вміння обирати змінні процесу, які можуть в достатній мірі характеризувати його статистику або динаміку, вміння збирати статистичні дані та використовувати їх для навчання мережі. Взагалі задача побудови мережі Баєса пов'язана з необхідністю послідовного розв'язання декількох підзадач обчислювального характеру, що зустрічаються при процесі навчання нейронної мережі. В загальному випадку навчання мережі Баєса відноситься до NP - повних задач. Це означає, що із збільшенням кількості вузлів в мережі, об'єм обчислень зростає поліноміально [16-17].

Дерева рішень - використовуються у машинному навчанні як передбачувальні моделі, що відображають знання про об'єкт у множину рішень. Дерева рішень є одним з підходів щодо передбачувального моделювання у статистиці, інтелектуального аналізу даних та машинного навчання. Модель є формою навчання під наглядом, що означає, що модель навчається та перевіряється на наборі даних, який містить бажану

категоризацію. Дерево рішень не завжди може давати чітку відповідь або рішення. Натомість він може надавати варіанти, щоб спеціаліст з даних міг самостійно прийняти обґрунтоване рішення. Древа рішень імітують людське мислення, тому дослідникам обробки даних зазвичай легко зрозуміти й інтерпретувати результати [16].

Дерево рішень - це деревоподібна структура, де кожен внутрішній вузол позначає функцію, гілки - правила, а листові вузли - результат алгоритму. Це універсальний алгоритм машинного навчання з керуванням, який використовується як для задач класифікації, так і для задач регресії. Це один з дуже потужних алгоритмів. Він також використовується в Random Forest для навчання на різних підмножинах навчальних даних, що робить Random Forest одним із найпотужніших алгоритмів у машинному навчанні [17].

З точки зору алгоритмістики, дерево рішень функціонує шляхом аналізу набору даних для прогнозування його класифікації. Аналіз починається з кореневого вузла дерева, де алгоритм переглядає значення кореневого атрибута в порівнянні з атрибутом запису в фактичному наборі даних. Виходячи з порівняння, він переходить до відгалуження і переходить до наступного вузла. Алгоритм повторює цю дію для кожного наступного вузла, порівнюючи його значення атрибутів з значеннями підвузлів і продовжуючи процес далі. Процес буде продовжуватись до тих пір, доки не буде досягнуто останнього листового вузла дерева [16].

Для побудови дерев рішень використовується алгоритм CART (Classification and Regression Tree). Він працює, вибираючи найкращий поділ на кожному вузлі на основі таких показників, як домішка Gini або інформаційний приріст [15, 17].

Сильними сторонами методів дерев рішень є:

- здатність генерувати зрозумілі правила;
- вміють виконувати класифікацію, не вимагаючи особливих обчислень;

- Дерева рішень здатні обробляти як безперервні, так і категорійні змінні;
- Дерева рішень дають чітке уявлення про те, які галузі є найбільш важливими для прогнозування або класифікації;
- Простота використання: Дерева рішень прості у використанні та не вимагають великих технічних знань, що робить їх доступними для широкого кола користувачів;
- Масштабованість: Дерева рішень можуть обробляти великі набори даних і можуть бути легко розпаралелені для скорочення часу обробки;
- Толерантність до відсутніх значень: Дерева рішень здатні обробляти відсутні значення в даних, що робить їх підходящим вибором для наборів даних з відсутніми або неповними даними;
- Обробка нелінійних зв'язків: Дерева рішень можуть обробляти нелінійні зв'язки між змінними, що робить їх підходящим вибором для складних наборів даних.
- Здатність обробляти незбалансовані дані: Дерева рішень можуть обробляти незбалансовані набори даних, де один клас сильно представлений порівняно з іншими, зважаючи важливість окремих вузлів на основі розподілу класів.

Серед слабких сторін методів дерева рішень можна відзначити наступні:

- Дерева рішень менш придатні для завдань оцінки, де метою є прогнозування значення неперервного атрибута;
- Дерева рішень схильні до помилок в задачах класифікації з багатьма класами і відносно невеликою кількістю навчальних прикладів;
- Навчання дерев рішень може бути обчислювально дорогим. Процес вирощування дерева рішень є обчислювально дорогим. У кожному вузлі кожне поле розбиття кандидата має бути відсортоване, перш ніж можна буде знайти його найкращий розподіл. У деяких алгоритмах

використовуються комбінації полів і здійснюється пошук оптимального комбінування ваг. Алгоритми обрізки також можуть бути дорогими, оскільки необхідно сформувати та порівняти багато піддерев-кандидатів;

- Дерева рішень схильні до переобладнання тренувальних даних, особливо коли дерево дуже глибоке або складне. Це може призвести до низької продуктивності нових, невидимих даних;
- Невеликі варіації в навчальних даних можуть призвести до генерації різних дерев рішень, що може бути проблемою при спробі порівняти або відтворити результати;
- Багато алгоритмів дерева рішень погано обробляють відсутні дані і вимагають приписування або видалення записів з відсутніми значеннями;
- Початкові критерії розбиття, що використовуються в алгоритмах дерева рішень, можуть призводити до упереджених дерев, особливо при роботі з незбалансованими наборами даних або рідкісними класами;
- Дерева рішень обмежені у своїй здатності представляти складні взаємозв'язки між змінними, особливо коли вони мають справу з нелінійними або інтерактивними ефектами;
- Дерева рішень можуть бути чутливими до масштабування вхідних ознак, особливо при використанні метрик на основі відстані або правил прийняття рішень, які покладаються на порівняння між значеннями [17].

Для побудови моделі передбачення згідно завдання поточної магістерської кваліфікаційної роботи, було прийняте рішення використовувати саме дерева рішень.

1.3 Огляд та аналіз аналогів

Існує багато методів та реалізацій моделей передбачення та прогнозування генерації електроенергії фото електростанціями з урахуванням різних погодних факторів, в тому числі сонячної активності. Їх точність безумовно залежить від багатьох факторів, а саме обраної моделі, самого об'єкту досліджень, його географічного положення тощо. Всі методи та реалізації які походять від них, мають свої переваги та недоліки.

Існують декілька порівняльних досліджень, які обговорюються в багатьох наукових статтях, пов'язаних із методами машинного навчання та deep learning.

У статтях М.Боузердома, А.Меллита та М.Раза [11] застосовувалися методи SVM, ANN для прогнозування виробленої сонячної енергії. У статті М.Абуела [12] був використаний метод ANN для прогнозу сонячної енергії.

У роботі Сі Чена [19] обговорюється використання ANN з вдосконаленими статистичними процедурами для прогнозування сонячної енергії. Метод ANN застосовувався для прогнозування сонячної системи в іншій роботі [20].

У роботі Р.Ваз та Б.Елсінга [21] використовувався метод ANN для передбачення сонячної енергії з використанням спостережень близьких фотовольтаїчних систем та метеорологічних змінних.

Ще одне дослідження, обговорене в роботі Я.Ліна та І.Копринської [22], надало прогноз на один день вперед для сонячної системи, використовуючи алгоритм видобування послідовностей разом з нейронними мережами на основі LSTM. Її результат був кращим, ніж у нейронної мережі зворотного поширення (BP) та лінійної регресії найменших квадратів.

У роботі [23] порівнювалися три глибокі нейронні мережі для прогнозування сонячної енергії та надавали поради щодо вибору найкращої

мережі для реального використання. Прогнозування поведінки послідовних даних вимагає використання методу ANN.

Експериментальні результати в дослідженні [24] показали, що продуктивність методу прогнозування генерації електроенергії на основі LSTM краща, ніж у багатошарових перцептронів (MLP) і глибоких згорткових мереж.

Автори в [25] запропонували гібридну модель глибокого навчання CNN-LSTM, яка використовує багатошаровий CNN для вилучення ознак і рівень LSTM для прогнозування, тим самим ефективно покращуючи ефект прогнозування LSTM.

В статті [26] запропоновано гібридну модель прогнозування на основі SSA-LSTM. Декомпозиція SSA виконується на заданій послідовності вихідної потужності PV з сильною волатильністю; витягується послідовність трендів, періодична послідовність і послідовність шумів послідовності вихідної потужності фотоелектричних модулів; і за послідовністю виконується аналіз головних компонент. Для реконструкції послідовностей витягуються важливі компоненти, а для реконструйованих послідовностей окремо встановлюються моделі прогнозування LSTM. Мета полягає в тому, щоб дозволити моделі LSTM безпосередньо вивчати дані регулярних послідовностей, зменшити її складність та підвищити точність прогнозування.

Всі вищезгадані дослідження та роботи на аналогічну тему мають відносно непогані результати. Кожна отримана модель має свої переваги, а також звичайно недоліки, а отже завжди існує необхідність покращення, вдосконалення або дослідження та розробки принципово нового підходу, методу або моделі. Саме тому, поточна робота присвячена дослідженню та розробці ряду моделей передбачення добового генерування електроенергії фотоелектростанцією з урахуванням впливу сонячної активності з використанням технологій машинного навчання, серед яких була вибрана оптимальна. Модель має хороший результат передбачення з досить

маленькою похибкою. Застосування такої моделі на практиці зможе дати значний економічний ефект для фотоелектричної станції.

1.4 Висновки

В першому розділі магістерської кваліфікаційної роботи здійснено аналіз предметної області, розглянуто наявність існуючих проблем, їх актуальність та шляхи вирішення. Досліджено основні методи, що використовуються в машинному навчанні для побудови моделей прогнозування та передбачення. Розглянуто аналоги, деякі з яких вже мали практичне застосування.

Аналіз показав, що існує необхідність в розробці нових моделей передбачення добового генерування електроенергії фотоелектростанцією з урахуванням впливу сонячної активності, використовуючи технології машинного навчання. Отримані моделі мають бути досліджені на предмет похибок передбачення, з відповідним супроводженням графічними діаграмами та графіками. Серед отриманих моделей має бути вибрана одна, яка забезпечує найкращий результат передбачення добового генерування електроенергії з найменшою похибкою.

2 ВИБІР ОПТИМАЛЬНИХ НАЛАШТУВАНЬ ДЛЯ РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ

2.1 Веб платформа Kaggle

Для розв'язку задачі, що поставлена в поточній магістерській кваліфікаційній роботі, використовується хмарне обчислювальне середовище Kaggle.

Kaggle - це веб-платформа, на якій розміщено найбільшу у світі спільноту Data Science, яка налічує понад 536 000 активних учасників у 194 країнах і отримує майже 150 000 заявок на місяць, а також надає їй потужні інструменти та ресурси, які допомагають досягти хороших результатів у галузі науки про дані. Kaggle, як і Datascientest, пропонує налаштоване середовище Jupyter Notebooks без конфігурації. Графічні процесори та велика кількість даних і коду, опублікованих спільнотою, знаходяться у вільному доступі. У Kaggle знаходиться багата кількість коду і даних, необхідних для виконання практично будь-яких проектів з Data Science. Існує понад 50 000 публічних наборів даних і 400 000 загальнодоступних блокнотів, доступних для використання кожним [27]. Головний екран платформи показано на рисунку 2.1.

Цій платформі довіряють деякі з найбільших компаній у світі Data Science, такі як Walmart або Facebook. Платформа дозволяє фахівцям з обробки даних та іншим розробникам брати участь у конкурсах машинного навчання або челенджах даних, писати та ділитися кодом, а також розміщувати набори даних.

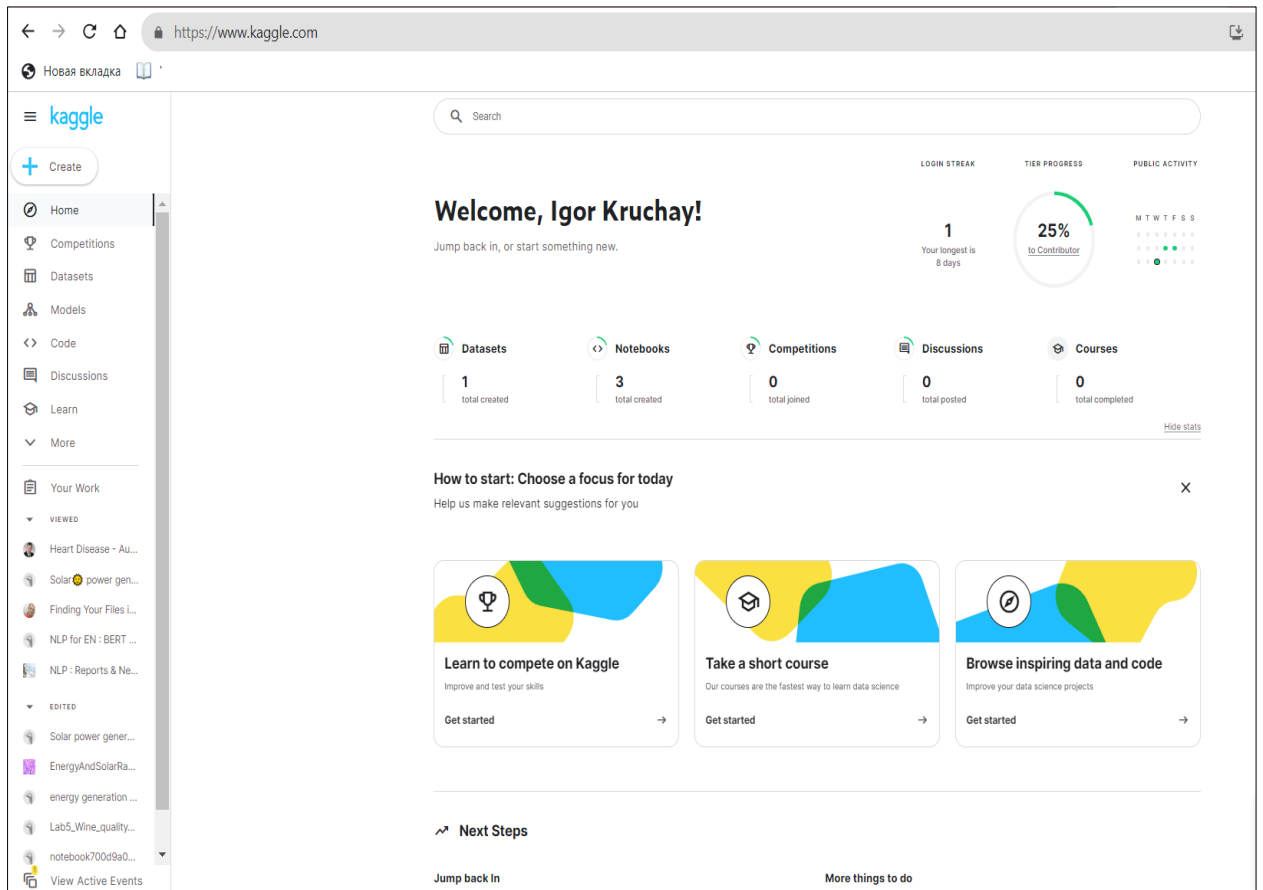


Рисунок 2.1 – Головний екран платформи Kaggle

Теми, які висвітлюються на Kaggle, надзвичайно різноманітні, починаючи від спроб передбачити початок раку шляхом вивчення карток пацієнтів і закінчуючи аналізом настроїв рецензій на фільми. Платформа пропонує цікаві та складні проекти, де користувачі можуть навчатися та практикуватися, а також пропонує дискусії з лідерами галузі та експертами.

Kaggle надає потужні ресурси в хмарі та дозволяє використовувати максимум 30 годин графічного процесора та 20 годин TPU на тиждень. У Kaggle можна завантажувати свої набори даних, а також набори даних інших користувачів.

Хоча в Kaggle офіційно підтримуються лише мови програмування Python і R, також можна створити будь-який компілятор із вихідних кодів і запустити його у своєму ядрі та використовувати результати у своєму тимчасовому каталозі безпосередньо з Python або R.

Kaggle прагне надати вам інструменти, необхідні для того, щоб стати дослідником даних світового рівня. Він надає доступ до реальних даних у режимі реального часу, щоб можна було попрактикуватися у вирішенні проблем, подібних до тих, з якими стикаються компанії в усьому світі [28].

2.2 Мова програмування Python

Для задачі, що поставлена в поточній магістерській кваліфікаційній роботі, використовується мова програмування Python.

Python - це мова програмування, яка широко використовується у веб-додатках, розробці програмного забезпечення, науці про дані та машинному навчанні (ML). Розробники використовують Python, оскільки він ефективний, легко вивчається і працює на різних платформах. Програми на мові Python можна завантажити безкоштовно, вони сумісні з усіма типами операційних систем і прискорюють швидкість розробки.

Мова програмування Python має декілька важливих переваг:

- **Читабельність коду:** Синтаксис Python досить простий для розуміння і нагадує англійську мову, що полегшує читання та розуміння коду;
- **Підвищення продуктивності:** Python дозволяє розробникам писати більш компактні програми, використовуючи менше рядків коду порівняно з деякими іншими мовами програмування, що сприяє підвищенню продуктивності;
- **Інтеграція з іншими мовами:** Розробники можуть комбінувати Python з іншими мовами програмування, такими популярними як Java, C і C++, що забезпечує гнучкість у розробці;
- **Велика стандартна бібліотека:** Python постачається з обширною стандартною бібліотекою, яка містить багато корисних модулів і інструментів, що спрощує виконання багатьох задач без необхідності написання коду з нуля;

- Активна спільнота: Python має активну, велику спільноту розробників, що забезпечує підтримку, обмін досвідом і вирішення виникаючих проблем;
- Багато навчальних ресурсів: Інтернет надає багато ресурсів для вивчення Python, включаючи відеоролики, навчальні посібники, документацію і керівництва для розробників;
- Переносимість: Python легко переноситься на різні операційні системи, такі як Windows, macOS, Linux і Unix.

Спеціалісти з обробки даних використовують бібліотеки Python ML для моделей машинного навчання і класифікаторів, які можуть точно класифікувати дані. Класифікатори, що побудовані на основі Python використовуються в різних областях і застосовуються для виконання завдань, таких як класифікація зображень, тексту та мережевого трафіку, розпізнавання мови та обличчя. Спеціалісти також використовують Python для глибокого навчання - передової техніки машинного навчання [29].

Більш конкретні завдання, які можуть вирішуватися за допомогою Python:

- Обробка та підготовка даних;
- виправлення та видалення нерелевантних даних;
- Вибір різних характеристик даних;
- Розмітка даних, що додає даним прості та зрозумілі імена;
- Пошук статистичної інформації в наборах даних;
- Візуалізація за допомогою діаграм та графіків: лінійних діаграм, стовпчатих діаграм, гістограм та кругових діаграм тощо.

В мові Python широко використовуються бібліотеки. Це набір кодів, які розробники можуть включати в свої програми, щоб не писати код з нуля. За замовчуванням у Python доступна стандартна бібліотека, яка містить велику кількість функцій, що багаторазово використовуються. Крім того, доступно

більше 137 000 бібліотек Python для різних завдань, серед яких інтернет-розробка, наука про дані та машинне навчання [30].

Перелік основних бібліотек Python:

- Pandas. Дана бібліотека містить гнучкі та оптимізовані структури даних. Ці структури можна використовувати для управління часовими рядами, структурованими даними, такими як таблиці, масиви, списки тощо. Для прикладу, бібліотеку Pandas можна використовувати для читання, запису, об'єднання, фільтрації та групування даних. Також бібліотека широко застосовується при аналізі даних та завданнях машинного навчання;

- Matplotlib. Бібліотека використовується для відображення даних у високоякісній двовимірній (2D) та тривимірній (3D) графіці. Matplotlib широко застосовують під час вирішення наукових завдань. За допомогою цієї бібліотеки дані можна візуалізувати у вигляді різних діаграм, таких як стовпчасті та лінійні. Можна будувати та відображати кілька діаграм одночасно, а графіку переносити на будь-які платформи;

- NumPy. Є популярною бібліотека, яка використовується для простого створення та управління масивами, логічними фігурами та виконання операцій, що відносяться до лінійної алгебри. Бібліотека підтримує інтеграцію з багатьма відомими мовами. Наприклад C++ та C;

- Keras. Це бібліотека для глибоких нейронних мереж Python з великою кількістю зручних функцій для обробки, візуалізації даних та багато чого іншого. Бібліотека підтримує безліч нейронних мереж. Keras має модульну структуру, що забезпечує неабияку гнучкість під час розробки інноваційних додатків;

- OpenCV-Python. Використовується для обробки зображень під час роботи з машинним зором. Бібліотека містить безліч складних в реалізації, але простих в користуванні функцій для обробки зображень та відеоінформації. Серед таких функцій можна виділити: перетворення

двовимірного середовища на тривимірне, одночасне читання та запис зображення, захват зображення з відео та його аналіз [29].

2.3 Дослідження та аналіз вхідних даних

Створюємо проект в платформі Kaggle. Далі, до проекту підключаються всі необхідні бібліотеки. Сюди відносяться: загальні бібліотеки Python, бібліотеки для моделювання, візуалізації, збору метрик тощо. Процес підключення бібліотек зображений на рисунку 2.2.

```
# Work with Data - the main Python libraries
import numpy as np
import pandas as pd
import pandas_profiling as pp

# Visualization
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

# Preprocessing
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split, KFold, ShuffleSplit, GridSearchCV

# Modeling
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
import xgboost as xgb
from xgboost.sklearn import XGBRegressor

# Metrics
from sklearn.metrics import r2_score

import warnings
warnings.simplefilter('ignore')
```

Рисунок 2.2 – Підключення бібліотек до проекту

Далі необхідно завантажити вхідний датасет (файл формату csv), в якому міститься погодинна інформація про об'єм генерації електричної енергії, а також показники сонячної активності (радіації) в цьому регіоні.

Початковий датасет, якій був переданий з фотоелектричної станції, не підходив в тому вигляді, в якому він був, отже його необхідно було привести до іншого, зручного для обробки вигляду. Для цього, за допомогою мови програмування Java, був розроблений конвертер даних, який на вхід бере початковий датасет, а на виході формує новий файл, в якому дані розташовані в іншому порядку. Крім цього, за допомогою конвертера, із початкового датасету були видалені інші погодні показники, які також прямо чи опосередковано впливають на генерацію електроенергії фотоелектричною станцією, оскільки вони не використовуються в даній роботі, але займають додаткову пам'ять та простір на диску.

Опис структури конвертера даних не входить в межі поточної магістерської кваліфікаційної роботи, оскільки є допоміжним інструментом. Отже, далі по тексту, мова буде йти про вже трансформований датасет, якій буде мати назву - вхідний.

Завантаження вхідного датасету зображено на рисунку 2.3.

```
df_solar = pd.read_csv(  
    '/kaggle/input/energyandsolarradiation/EnergyAndSolarRadiation.csv',  
    sep=';',  
    header=0  
)
```

Рисунок 2.3 – Завантаження вхідного датасету

Для завантаження використовується стандартна функція бібліотеки Pandas - read_csv. На вхід функції передається шлях до файлу з вхідним

датасетом, символ що є розділювачем даних та рядок файлу, якій містить заголовок.

Для виводу датасету на екран, вказується його назва. На рисунку 2.4 показано завантажений та виведений на екран датасет в табличному вигляді з наступними стовпцями:

«DATE_TIME» – дата та час фіксації показників;

«ENERGY» – об’єм генерації електроенергії [кВт*год];

«SOLAR_RADIATION» – значення зафіксованої сонячної радіації [Вт/м²].

	DATE_TIME	ENERGY	SOLAR_RADIATION
0	22.02.2020 7:30	46	49
1	22.02.2020 8:00	93	102
2	22.02.2020 8:30	152	151
3	22.02.2020 9:00	426	467
4	22.02.2020 9:30	406	455
...
16168	23.02.2022 15:00	141	143
16169	23.02.2022 15:30	77	85
16170	23.02.2022 16:00	45	48
16171	23.02.2022 16:30	45	50
16172	23.02.2022 17:00	7	7

16173 rows × 3 columns

Рисунок 2.4 – Датасет в табличному вигляді

Використовуючи функцію - info, можна відобразити інформацію про датасет: кількість стрічок з даними, кількість стовпців з їх назвами та кількість

пам'яті, яка використовується для завантаженого датасета. На рисунку 2.5 показано використання функції `info`, та результат її роботи.

```
df_solar.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16173 entries, 0 to 16172
Data columns (total 3 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   DATE_TIME              16173 non-null  object
1   ENERGY                16173 non-null  int64
2   SOLAR_RADIATION        16173 non-null  int64
dtypes: int64(2), object(1)
memory usage: 379.2+ KB
```

Рисунок 2.5 – Інформація про датасет

Для зручності аналізу вхідних даних, в завантажений датасет необхідно додати нові стовпці, які будуть містити нові згенеровані дані, що є результатом обробки даних існуючих. До нових стовпців відносяться:

«DATE» - дата вимірювання;

«TIME» - час вимірювання;

«YEAR» - окремо рік вимірювання;

«MONTH» - окремо місяць вимірювання;

«DAY» - окремо день вимірювання;

«TOTAL_TIME» - кількість хвилин, що пройшло від першого вимірювання;

«TOTAL_DAYS» - кількість днів, що пройшло від першого вимірювання;

«DATE_AS_STRING» - дата трансформована в стрічковий тип, яка використовується для побудови графіків.

На рисунку 2.6 показано фрагмент коду для розширення датасету.

```
df_solar['DATE_TIME'] = pd.to_datetime(df_solar['DATE_TIME'], format = '%d.%m.%Y %H:%M')
```

+ Code + Markdown

```
# adding date and time columns
df_solar["DATE"] = pd.to_datetime(df_solar["DATE_TIME"]).dt.date
df_solar["TIME"] = pd.to_datetime(df_solar["DATE_TIME"]).dt.time

# adding date and time separate columns
df_solar["YEAR"] = pd.to_datetime(df_solar["DATE_TIME"]).dt.year
df_solar["MONTH"] = pd.to_datetime(df_solar["DATE_TIME"]).dt.month
df_solar["DAY"] = pd.to_datetime(df_solar["DATE_TIME"]).dt.day

# adding total values
time = pd.to_datetime(df_solar["TIME"], format='%H:%M:%S')
df_solar['TOTAL_TIME'] = time.dt.hour * 60 + time.dt.minute # total time in minutes
df_solar['TOTAL_DAYS'] = pd.to_datetime(df_solar["DATE_TIME"]).dt.dayofyear # total time in days

# adding date as string column
df_solar["DATE_AS_STRING"] = df_solar["DATE"].astype(str)
```

Рисунок 2.6 – Фрагмент коду для розширення датасету

Розширений датасет із новими доданими стовпцями показано на рисунку 2.7.

	DATE_TIME	ENERGY	SOLAR_RADIATION	DATE	TIME	YEAR	MONTH	DAY	TOTAL_TIME	TOTAL_DAYS	DATE_AS_STRING
0	2020-02-22 07:30:00	46	49	2020-02-22	07:30:00	2020	2	22	450	53	2020-02-22
1	2020-02-22 08:00:00	93	102	2020-02-22	08:00:00	2020	2	22	480	53	2020-02-22
2	2020-02-22 08:30:00	152	151	2020-02-22	08:30:00	2020	2	22	510	53	2020-02-22
3	2020-02-22 09:00:00	426	467	2020-02-22	09:00:00	2020	2	22	540	53	2020-02-22
4	2020-02-22 09:30:00	406	455	2020-02-22	09:30:00	2020	2	22	570	53	2020-02-22
5	2020-02-22 10:00:00	435	488	2020-02-22	10:00:00	2020	2	22	600	53	2020-02-22
6	2020-02-22 10:30:00	450	508	2020-02-22	10:30:00	2020	2	22	630	53	2020-02-22
7	2020-02-22 11:00:00	797	832	2020-02-22	11:00:00	2020	2	22	660	53	2020-02-22
8	2020-02-22 11:30:00	323	324	2020-02-22	11:30:00	2020	2	22	690	53	2020-02-22
9	2020-02-22 12:00:00	366	359	2020-02-22	12:00:00	2020	2	22	720	53	2020-02-22
10	2020-02-22 12:30:00	823	772	2020-02-22	12:30:00	2020	2	22	750	53	2020-02-22
11	2020-02-22 13:00:00	272	254	2020-02-22	13:00:00	2020	2	22	780	53	2020-02-22
12	2020-02-22 13:30:00	181	174	2020-02-22	13:30:00	2020	2	22	810	53	2020-02-22
13	2020-02-22 14:00:00	105	114	2020-02-22	14:00:00	2020	2	22	840	53	2020-02-22
14	2020-02-22 14:30:00	105	97	2020-02-22	14:30:00	2020	2	22	870	53	2020-02-22
15	2020-02-22 15:00:00	95	93	2020-02-22	15:00:00	2020	2	22	900	53	2020-02-22
16	2020-02-22 15:30:00	128	119	2020-02-22	15:30:00	2020	2	22	930	53	2020-02-22
17	2020-02-22 16:00:00	130	123	2020-02-22	16:00:00	2020	2	22	960	53	2020-02-22

Рисунок 2.7 – Розширений датасет

Обробка датасету закінчена, він підготовлений, тому можна розпочати аналіз вхідних даних.

Датасет містить дані за декілька років. Якщо вивести кількість вимірювань по роках, можна побачити, що найбільша кількість даних є для 2021 та 2020 років. Трохи більше ніж 7000 записів. За 2022 рік – найменша кількість записів, їх біля 1000. Розподілення показано на рисунку 2.8.

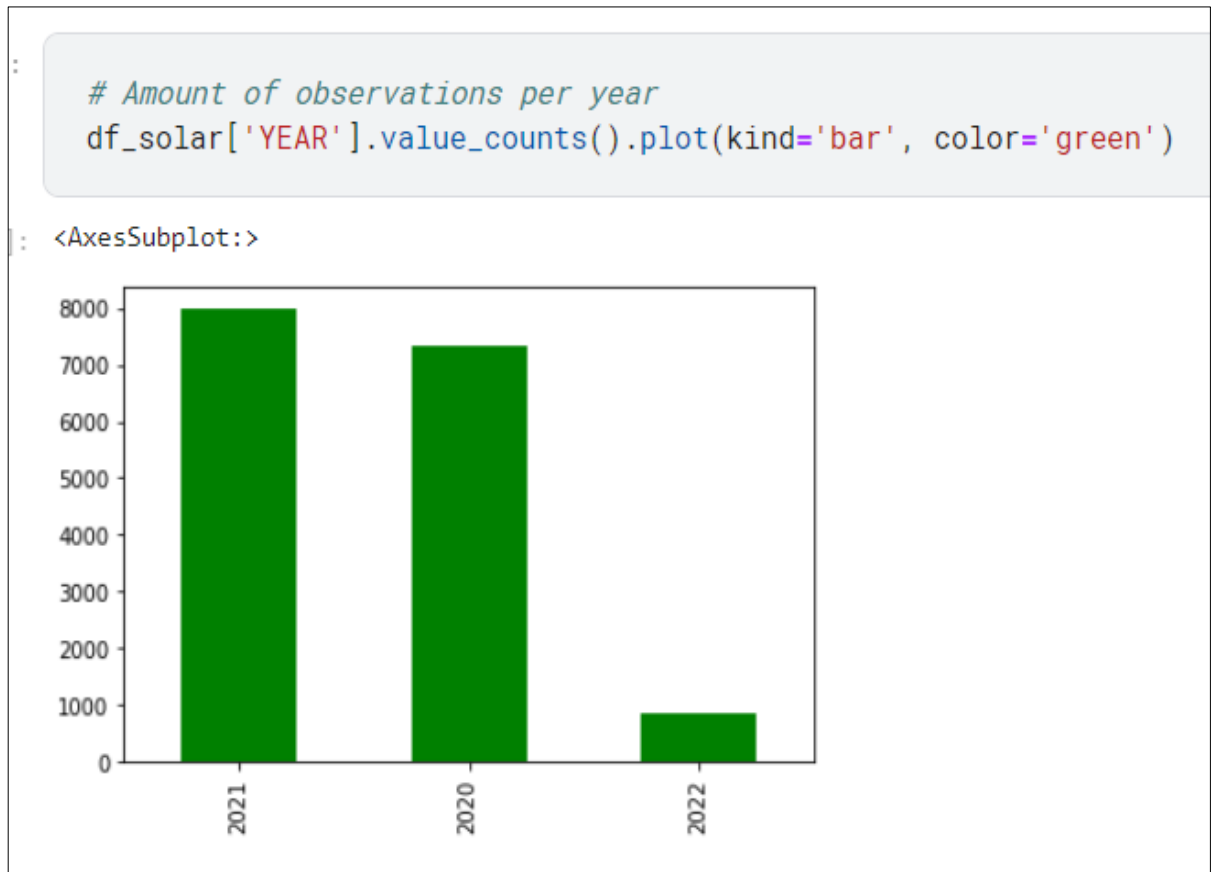


Рисунок 2.8 – Розподілення даних по роках

Побудуємо добові графіки генерації електроенергії за певний період часу 2020 року. Фрагмент коду, який призначений для цього, приведено на рисунку 2.9.


```

solar_dc = df_solar.pivot_table(values='ENERGY', index='TIME', columns='DATE')

def Daywise_plot(data= None, row = None, col = None, title='Energy generation', skip_day = None):
    cols = data.columns # take all columns
    gp = plt.figure(figsize=(20,40))

    gp.subplots_adjust(wspace=0.2, hspace=0.5)
    for i in range(1, col*row+1):
        ax = gp.add_subplot(row,col, i)
        data[cols[skip_day + i-1]].plot(ax=ax, color='blue')
        ax.set_title('{} {}'.format(title, cols[skip_day + i-1]),color='black')

Daywise_plot(data=solar_dc, row=12, col=3, skip_day = 30)

```

Рисунок 2.9 – Фрагмент коду для відображення добових графіків генерації електроенергії

В коді використовується функція `pivot_table` для агрегування даних з подальшим виводом їх у графічній формі за допомогою функцій `figure` і `subplot_adjust` бібліотеки `matplotlib`.

Результат побудови графіків генерації електроенергії за період від 26.03.20 по 09.04.20, відображений на рисунку 2.10.

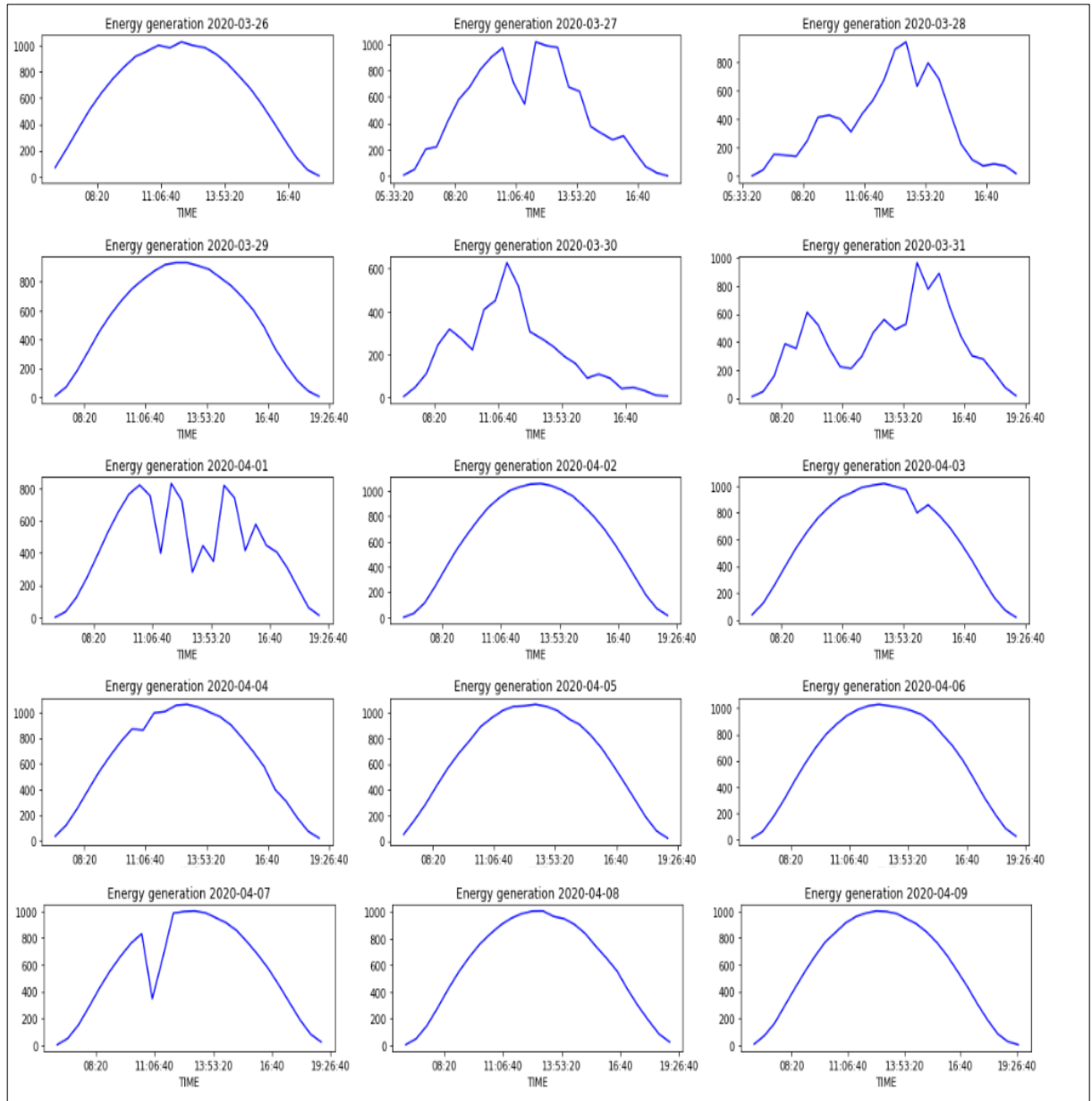


Рисунок 2.10 – Добові графіки генерації електроенергії

Побудуємо графіки сонячної активності в цей же самий період часу, щоб візуально побачити кореляцію між цими двома показниками. Фрагмент коду, який призначений для цього, наведено на рисунку 2.11.

```

solar_radiation = df_solar.pivot_table(values='SOLAR_RADIATION', index='TIME', columns='DATE')

def Daywise_plot(data= None, row = None, col = None, title='Solar radiation', skip_day = None):
    cols = data.columns # take all column
    gp = plt.figure(figsize=(20,40))

    gp.subplots_adjust(wspace=0.2, hspace=0.5)
    for i in range(1, col*row+1):
        ax = gp.add_subplot(row,col, i)
        data[cols[skip_day + i-1]].plot(ax=ax, color='red')
        ax.set_title('{} {}'.format(title, cols[skip_day + i-1]),color='black')

Daywise_plot(data=solar_radiation, row=12, col=3, skip_day = 30)

```

Рисунок 2.11 – Фрагмент коду для відображення добових графіків сонячної активності

Побудовані графіки сонячної за період від 26.03.20 по 09.04.20, відображені на рисунку 2.12.

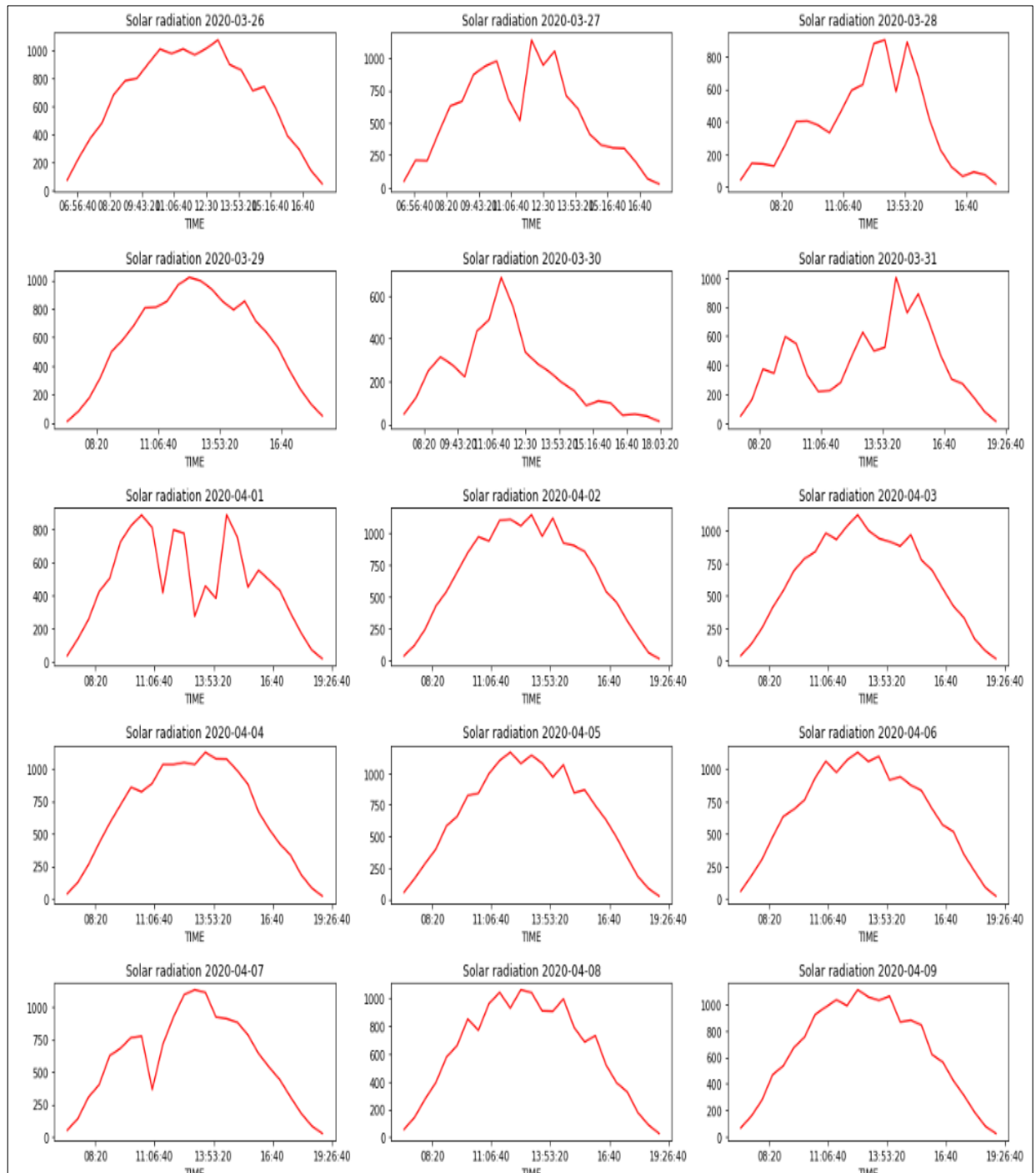


Рисунок 2.12 – Добові графіки сонячної активності

З графіків генерації електроенергії та сонячної активності добре видно, що їх форми в один і той самий час – дуже схожі. Отже можна стверджувати, що існує пряма залежність згенерованої електричної енергії від сонячної активності в цій місцевості в цей самий час.

На рисунках 2.13 та 2.14 показані об'єднані графіки за одну добу, де добре видно кореляцію між двома параметрами.

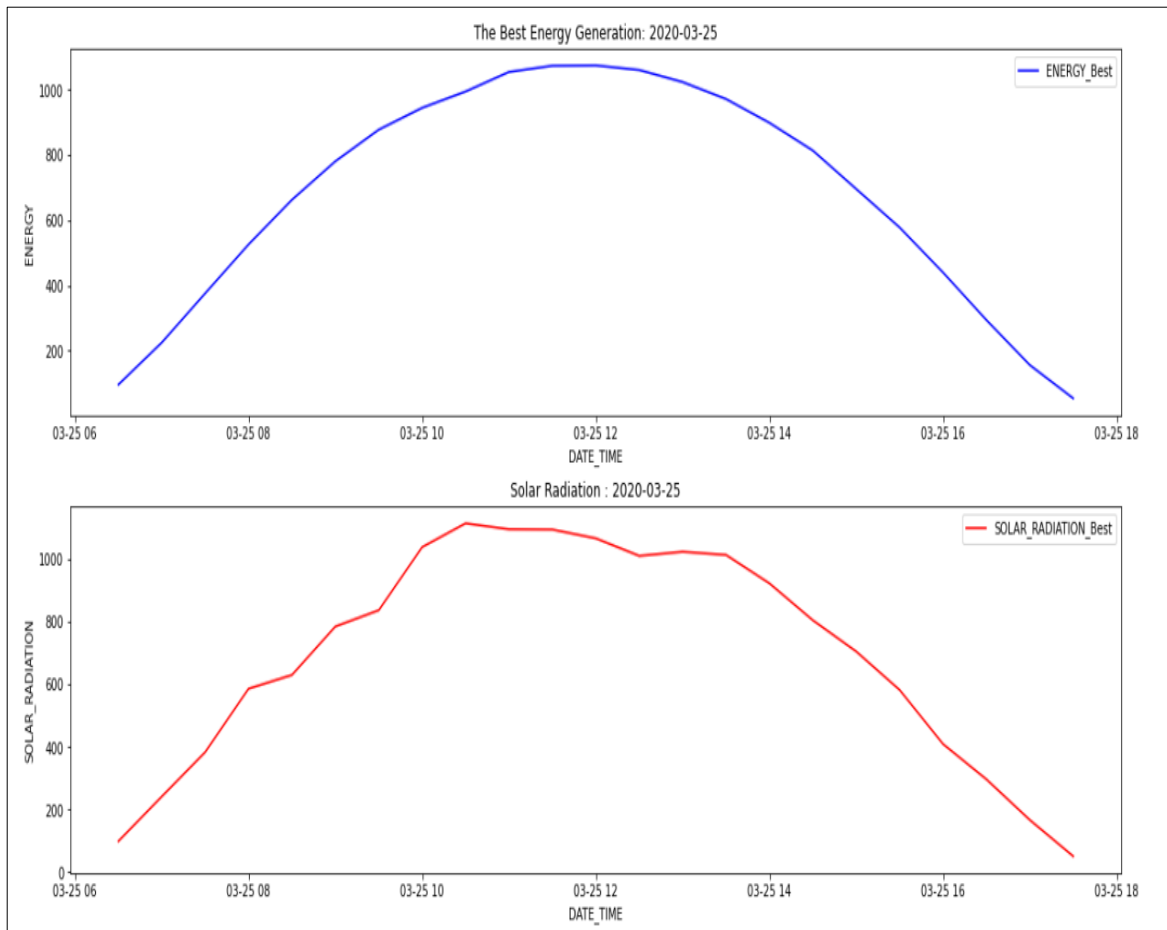


Рисунок 2.13 – Графіки генерації електроенергії та фіксації сонячної активності 25 березня 2020 року

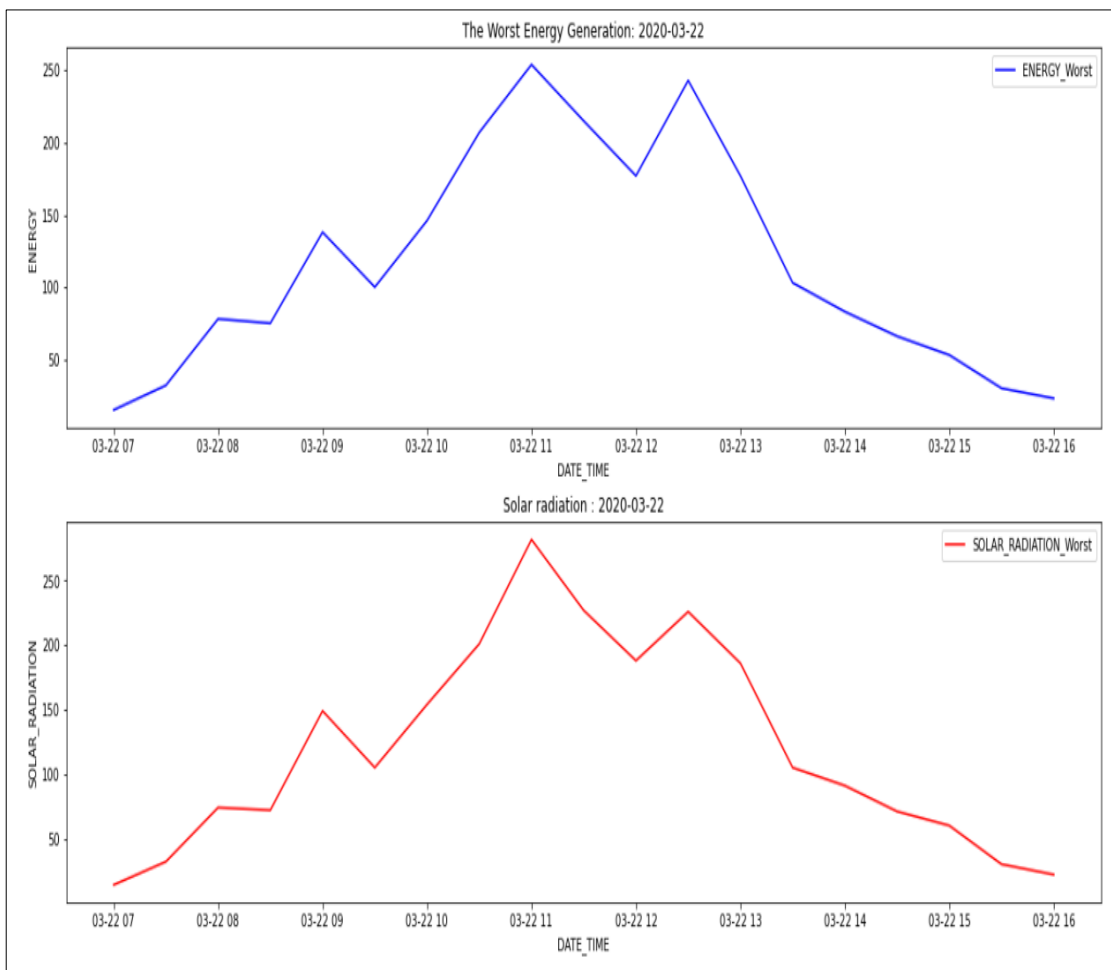


Рисунок 2.14 – Графіки генерації електроенергії та фіксації сонячної активності 22 березня 2020 року

Далі вхідні дані необхідно дослідити на предмет наявності аномальних значень, які погано впливають на кінцевий результат передбачення. Аномальними вважаються значення, які мають дуже суттєві відхилення від нормальних значень та суперечать суті самих даних. Для цього побудуємо погодинний графік генерації електроенергії, згрупувавши дані по «TIME», де значення енергії «ENERGY» вибираються за максимальним значенням. Фрагмент коду та результат наведено на рисунку 2.15.

```

daily_dc = df_solar.groupby('TIME')['ENERGY'].agg('max')

ax = daily_dc.plot.bar(figsize=(20,5), legend=True,color='blue')
plt.title('Hourly Maximum Energy Generation')
plt.show()

```

Рисунок 2.15 – Фрагмент коду для пошуку та відображення аномалій у вигляді стовпчастої діаграми

На рисунку 2.16 приведено стовпчасту діаграму, на якій добре видно, як аномальні значення відрізняються від нормальних.

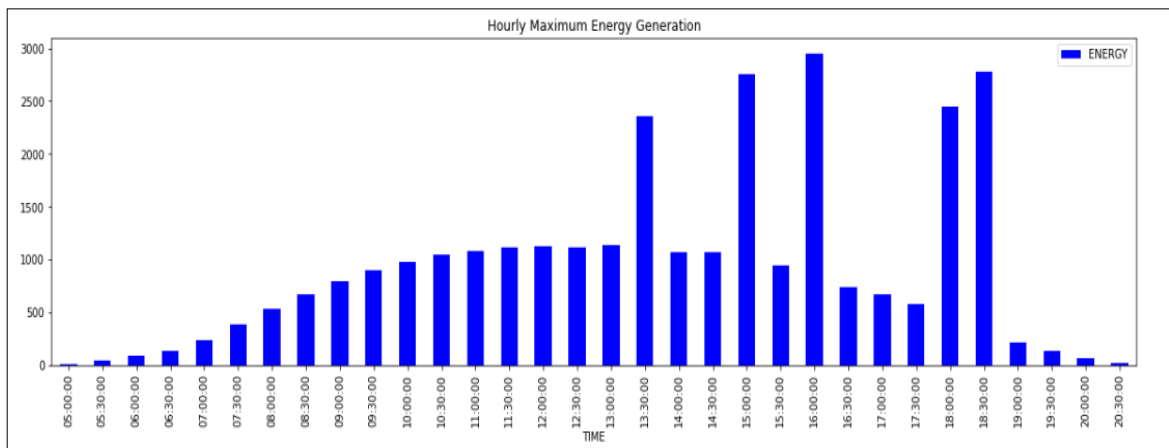


Рисунок 2.16 – Фрагмент коду для пошуку аномалій та результат його роботи у вигляді стовпчастої діаграми

Виходячи з того, що встановлена потужність фотоелектростанції становить 2.4 МВт, сумарна генерація за встановлений проміжок часу 30 хв, не може перевищувати 1200 кВт*год, тому значення, що перевищують зазначену величину будемо вважати аномальними. Значення генерації менші ніж 10 кВт*год [31], також вважаємо аномальними та будемо видаляти.

На рисунку 2.17 наведено фрагмент коду для видалення аномальних значень із датасету. Для цього використовується функція `loc`, з предикатом вибірки нормальних значень. Функція вибирає лише ті значення, які

задовольняють умові предиката.

```
df_solar = df_solar.loc[(df_solar["ENERGY"] > 10) & (df_solar["ENERGY"] < 1200)]
df_solar.describe()
```

Рисунок 2.17 – Видалення аномальних значень

На рисунках 2.14 та 2.15 показані відповідно максимальна місячна та загальна місячна генерація електроенергії. Можна побачити, що найбільша генерація електроенергії припадає саме на весняні та літні місяці, оскільки це обумовлено в тому числі і найбільшою кількістю сонячної радіації, яка попадає на земну поверхню. Що в свою чергу обумовлено нахилом останньої відносно сонця.

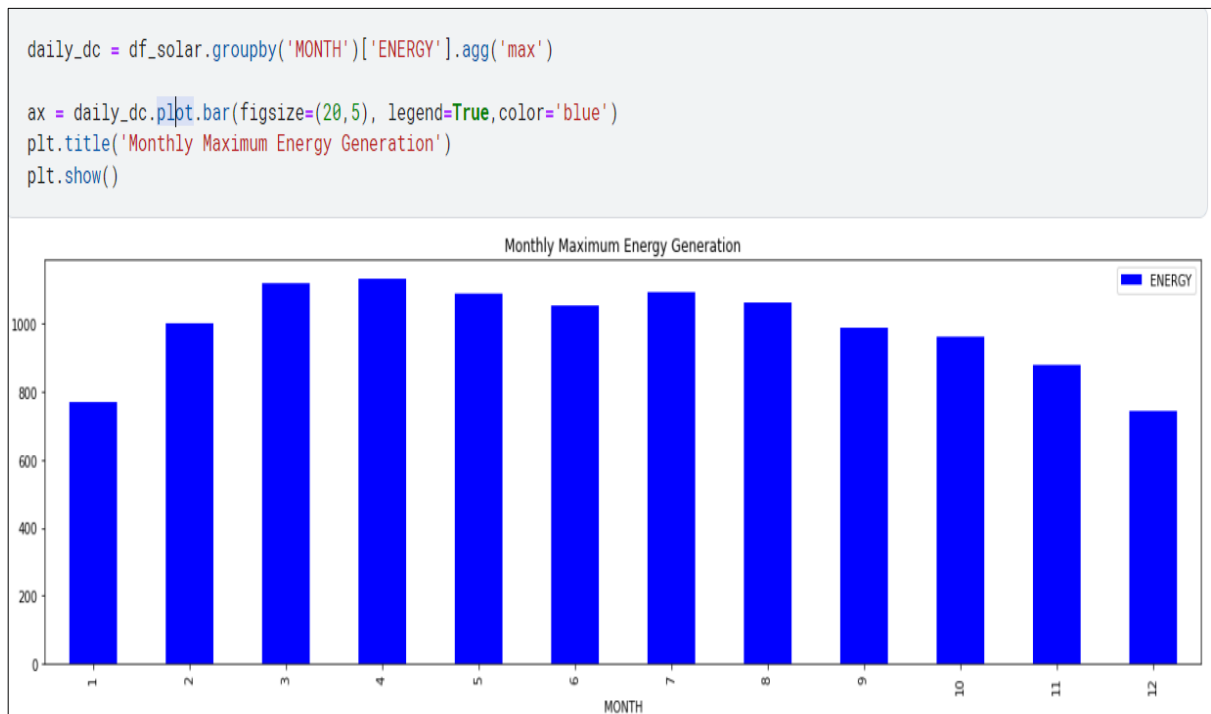


Рисунок 2.18 – Фрагмент коду та його результат для відображення максимальної місячної генерації електроенергії

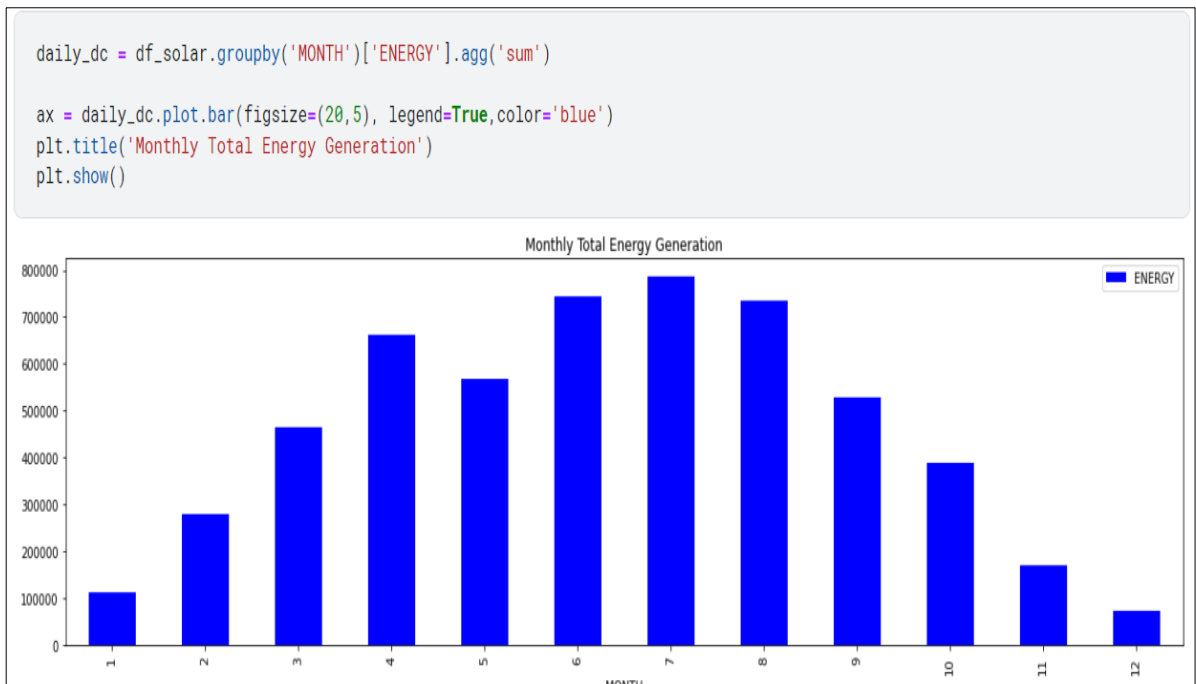


Рисунок 2.19 – Фрагмент коду та його результат для відображення загальної місячної генерації електроенергії

Проаналізувавши наведені графіки можна стверджувати, що кількість електроенергії, яка генерується сонячними панелями електростанції, напряму залежить від сонячної радіації, яка змінюється в залежності від сезону.

2.4 Розвідувальний аналіз даних та вибір оптимальних налаштувань

Розвідувальний аналіз - це попередній аналіз, який виконується з метою виявлення основних властивостей даних, законів розподілу та тенденцій, знаходження найбільш загальних закономірностей між ними, виявлення аномалій та побудови початкових моделей [32]. Результатами розвідувального аналізу зазвичай є діаграми, графіки, таблиці, схеми тощо. Далі перелічені методи, за допомогою яких відбувається розвідувальний аналіз:

- кластерний аналіз;

- аналіз часових рядів;
- дерева класифікацій;
- аналіз відповідностей;
- факторний аналіз;
- логлінійний аналіз;
- покрокова лінійна та нелінійна регресія;
- багатовимірне шкалювання;
- аналіз дискримінантних функцій;
- канонічні кореляції.

Як правило результатом розвідувального аналізу не користуються для вироблення управлінських рішень. Його застосування це допомога в розробці найкращої стратегії для аналізу, поглибленого вивчення, висування гіпотез чи інших математичних методів [32, 33].

Застосуємо один з інструментів бібліотеки Pandas – ProfileReport, для побудови інтерактивних звітів у вигляді різноманітних графіків та гістограм по кожному показнику [34] (рис. 2.20).

```
# EDA with Pandas Profiling
pp.ProfileReport(df_solar)
```

Рисунок 2.20 – Фрагмент коду із застосуванням інструменту ProfileReport

Після виконання команди на екран буде виведено інтерактивний статистичний звіт розділений на секції по показниках (стовпцям) датасету. На рисунку 2.21 відображена секція, що відповідає генерації електроенергії. На звіті видно основні статистичні показники: максимальне та мінімальне значення генерації, середнє та сумарне, показники 5-го та 95-го перцентілей, абсолютне та медіанне відхилення, а також інші.

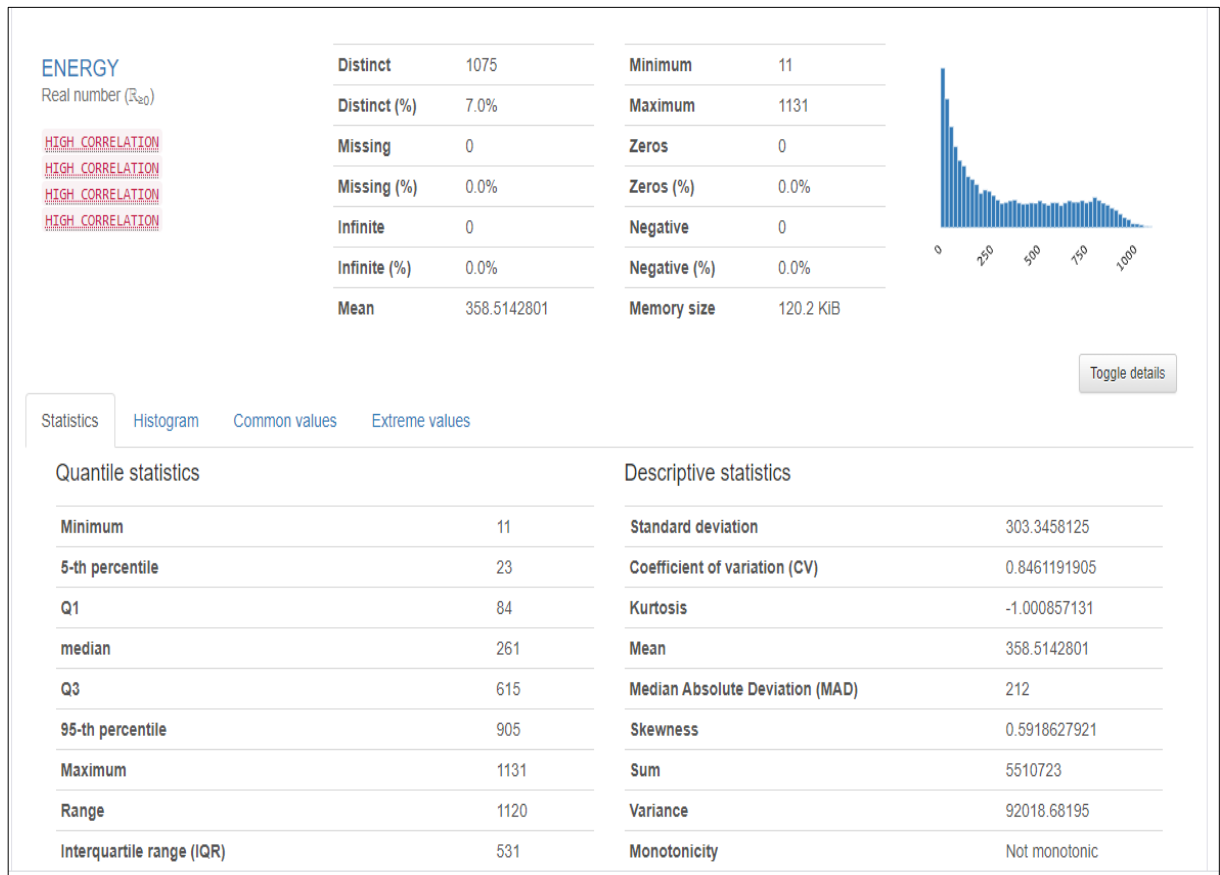


Рисунок 2.21 – Інтерактивний статистичний звіт щодо генерації електроенергії

На рисунку 2.22 відображена секція звіту, що відповідає сонячній активності. Так само, на звіті видно основні статистичні показники: максимальне та мінімальне значення сонячної активності, середнє та сумарне, показники 5-го та 95-го перцентілей, абсолютне та медіанне відхилення, а також інші.

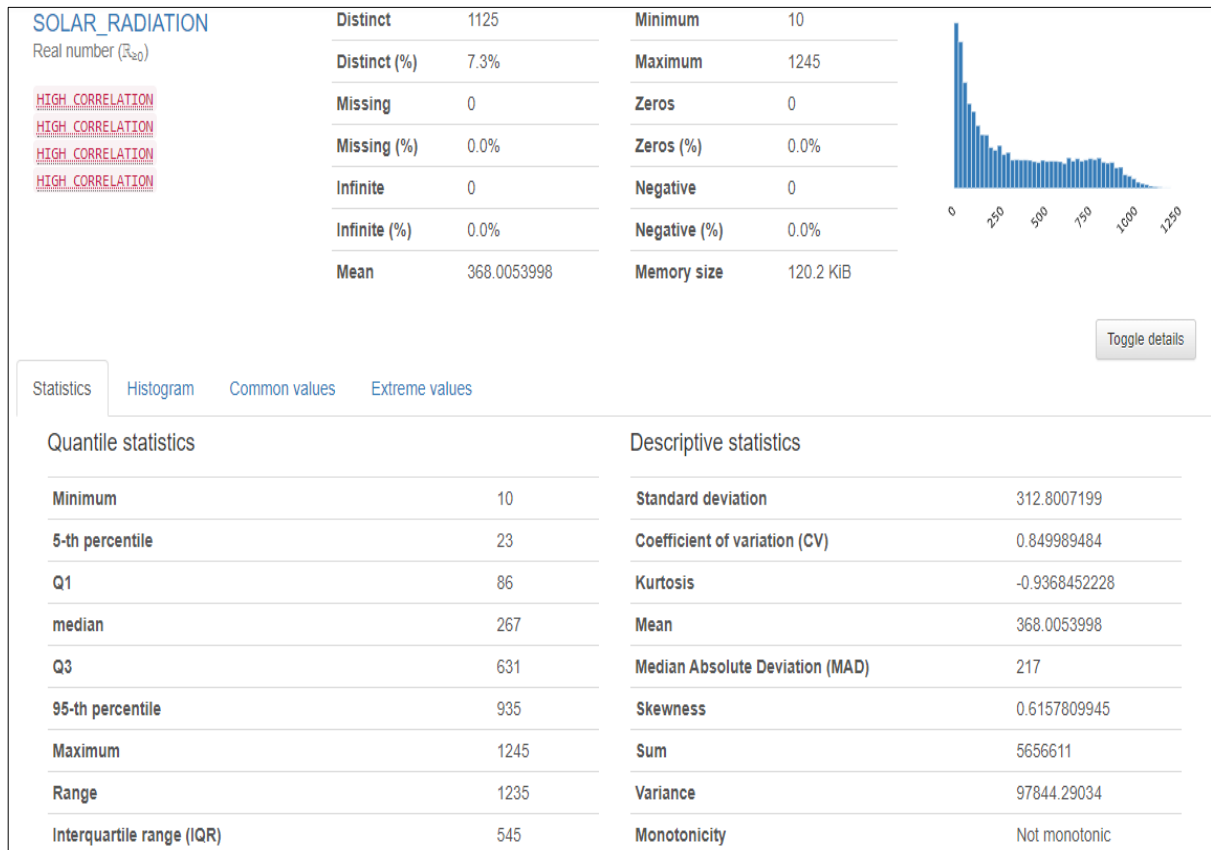


Рисунок 2.22 – Інтерактивний статистичний звіт щодо сонячної активності

Однією із цікавих особливостей «Profiling Report» є вкладка «Interactions» в якій автоматично створюються діаграми взаємодії для кожної пари показників датасету, що візуально відображає залежності між ними. Можна отримати діаграму взаємодії будь-якої пари, вибравши конкретні показники в заголовку. За допомогою таких діаграм можна оцінювати аномалії, наскільки вони близькі до лінійної регресії. В такий спосіб можна зробити попереднє вивчення співвідношень між даними показників [35].

На рисунку 2.23, відображена залежність між сонячною активністю та кількістю електричної енергії, що вироблена фотоелектростанцією.

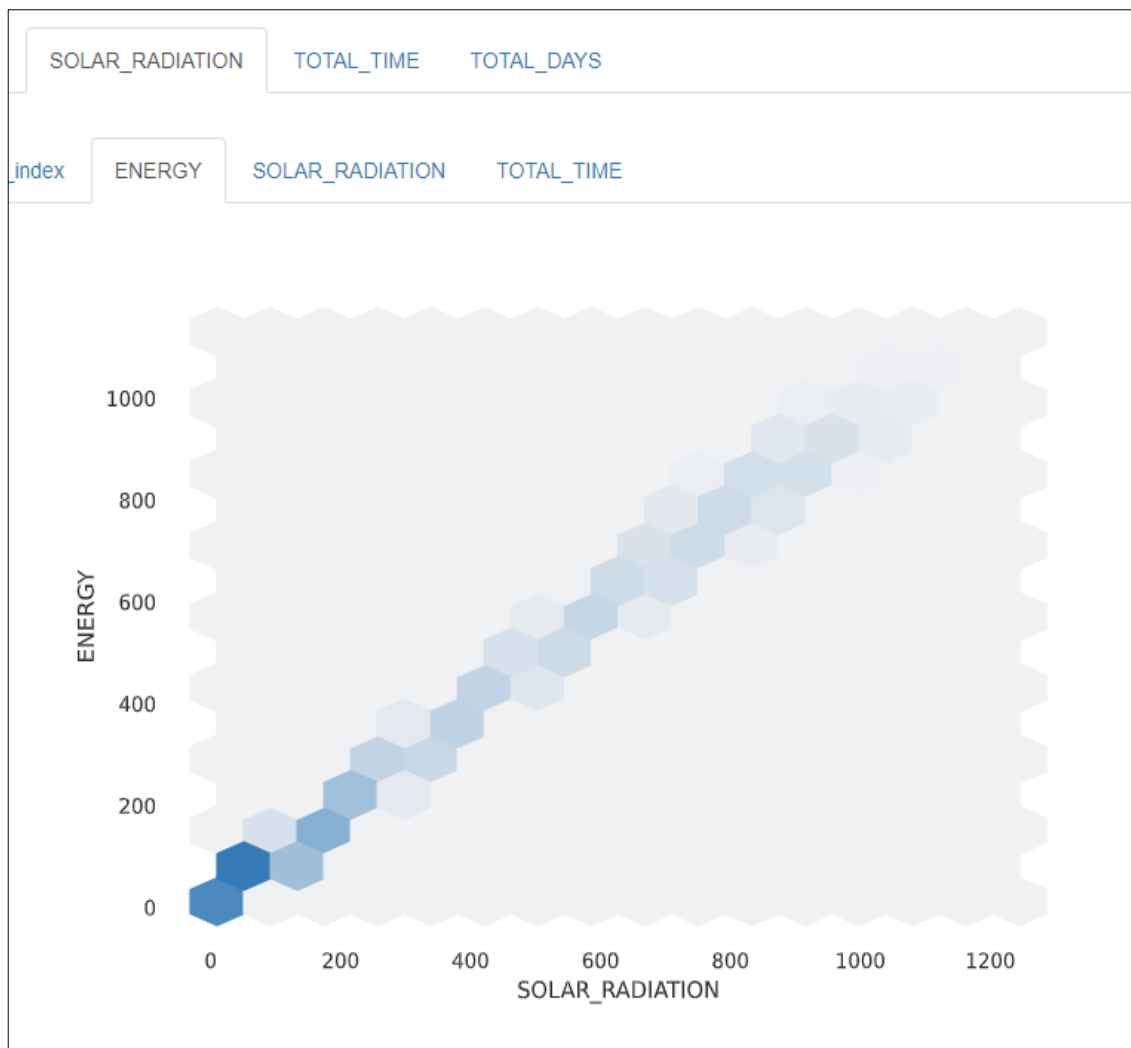


Рисунок 2.23 – Залежність між «ENERGY» та «SOLAR_RADIATION»

Ще однією особливістю «Profiling Report» є вкладка «Correlations», яка дозволяє побачити кореляцію між усіма показниками датасету, чи пов'язані вони між собою і наскільки сильно. «Profiling Report» дає можливість побудувати візуальні матриці: Пірсона, Спірмана, Кендалла та Phik. Далі розглянемо кожен з них [36-37].

Коефіцієнт кореляції Пірсона (r) є мірою лінійної кореляції між двома змінними. Значення коефіцієнту може бути в діапазоні $-1 \dots +1$. Значення -1 , вказує на загальну негативну лінійну кореляцію. Значення 0 , вказує на

відсутність лінійної кореляції. Значення 1, вказує на загальну позитивну лінійну кореляцію [38]. Крім того, кореляція інваріантна при окремих змінах розташування та масштабу двох змінних, маючи на увазі, що для лінійної функції кут до осі x не впливає на коефіцієнт кореляції. Матриця Пірсона, зображена на рисунку 2.24.

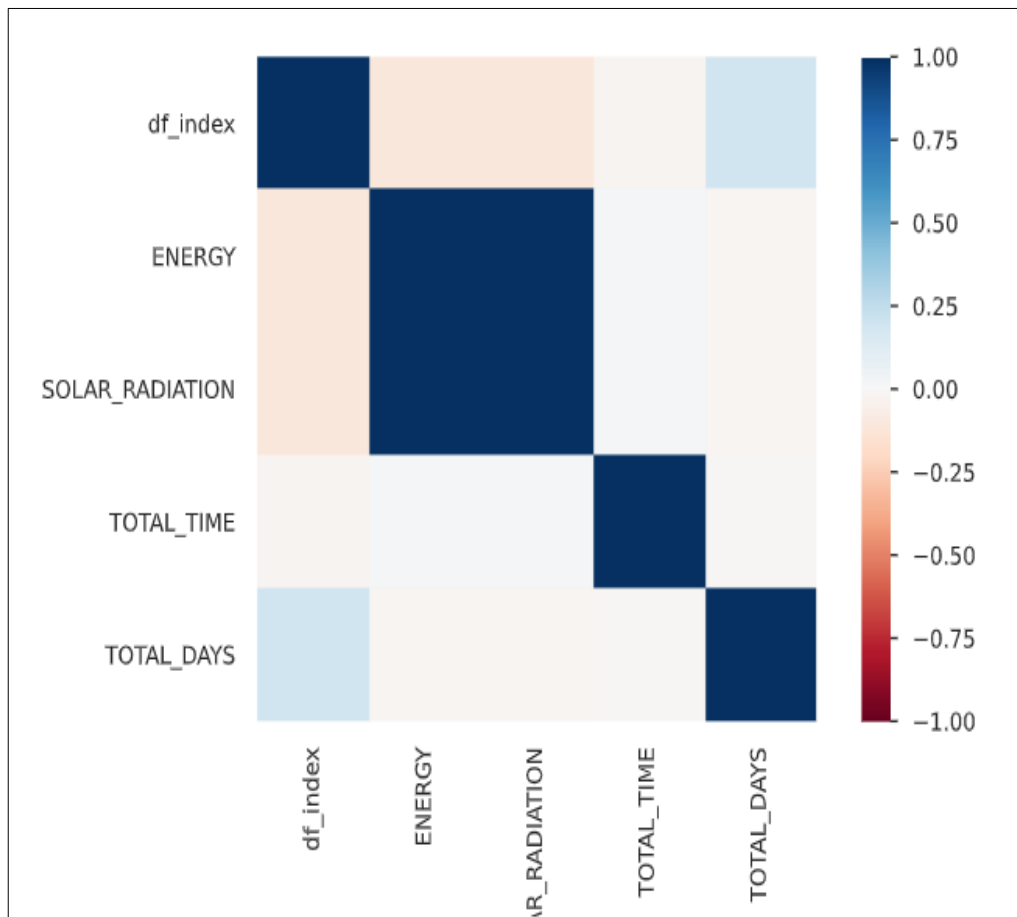


Рисунок 2.24 – Матриця кореляції Пірсона

Коефіцієнт рангової кореляції Спірмена (ρ) є мірою монотонної кореляції між двома змінними і тому є кращим серед нелінійних монотонних кореляцій, ніж коефіцієнт кореляції Пірсона. Взагалі кореляція Спірмена є менш чутливою, ніж кореляція Пірсона відносно сильних викидів. Значення коефіцієнту може бути в діапазоні $-1 \dots +1$. Значення -1 , вказує на загальну негативну монотонну кореляцію. Значення 0 , вказує на відсутність монотонної

кореляції. Значення 1, вказує на загальну позитивну монотонну кореляцію [39]. Матриця Спірмена зображена на рисунку 2.25.

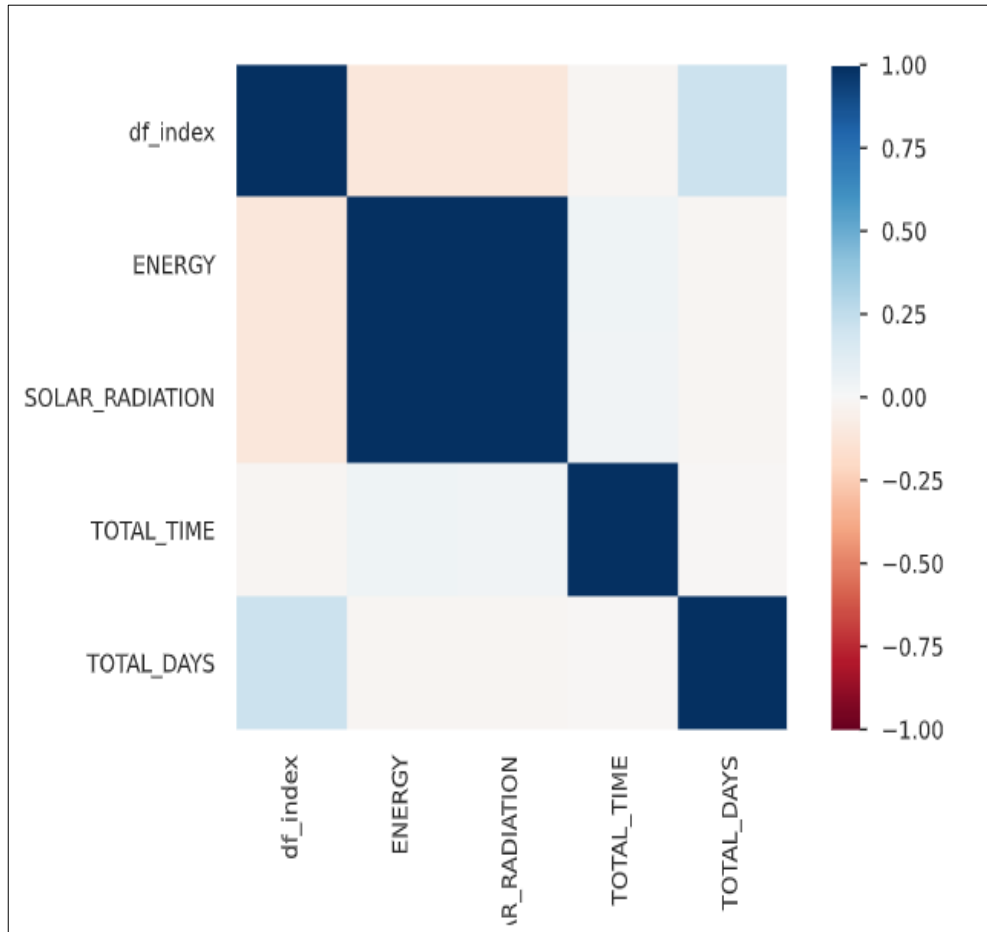


Рисунок 2.25 – Матриця кореляції Спірмена

Подібно до коефіцієнта рангової кореляції Спірмена, коефіцієнт кореляції рангу Кендалла (τ) вимірює порядковий зв'язок між двома змінними. Це значення також лежить в діапазоні $-1 \dots +1$. Значення -1 , вказує на загальну негативну кореляцію. Значення 0 , вказує на відсутність кореляції. Значення 1 , вказує на загальну позитивну кореляцію [40]. Матриця Кендалла зображена на рисунку 2.26.

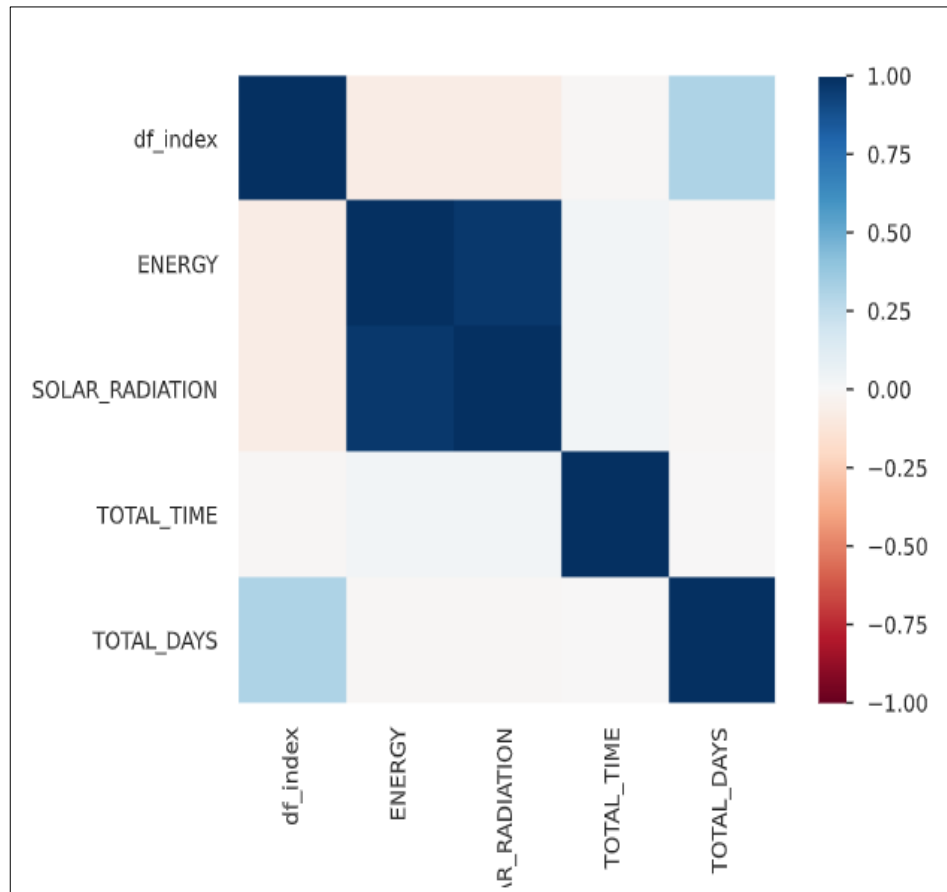


Рисунок 2.26 – Матриця кореляції Кендалла

Φ_k (ϕ_k) - це новий і практичний коефіцієнт кореляції, який працює послідовно між категоріальними, порядковими та інтервальними змінними, фіксує нелінійну залежність і повертається до коефіцієнта кореляції Пірсона у разі двовимірного нормального розподілу вхідних даних [41]. Матриця кореляції Φ_k зображена на рисунку 2.27.

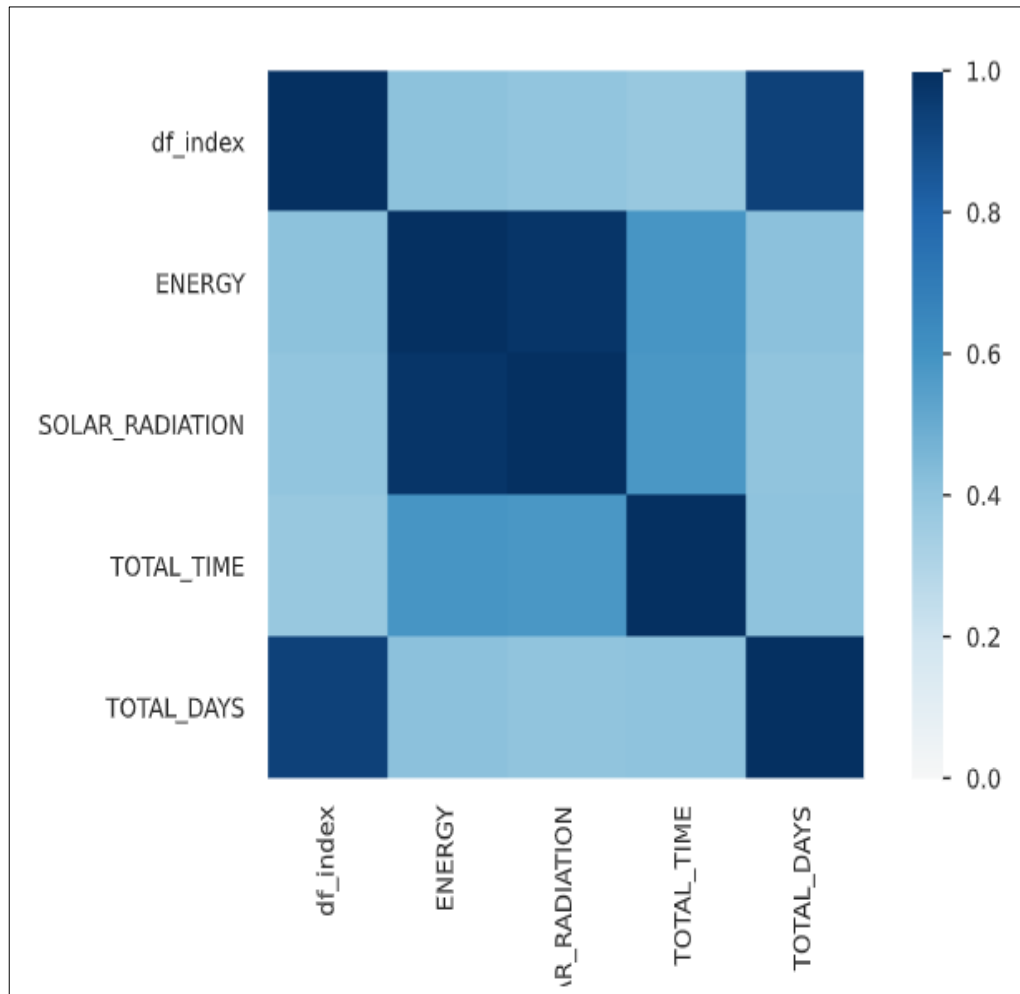


Рисунок 2.27 – Матриця кореляції ρ_{ik}

Аналізуючи коефіцієнти кореляції на матрицях Пірсона, Спірмена, Кендалла і ρ_{ik} , ми можемо бачити, що генерація електроенергії та показник сонячної активності корелюють між собою приблизно в середньому на 93%.

2.5 Висновки

В другому розділі магістерської кваліфікаційної роботи здійснено аналіз вхідних даних та їх попередня обробка. Були видаленні аномальні значення. Побудовані та проаналізовані добові графіки генерації електроенергії та сонячної радіації, визначена наявність кореляції між цими показниками. В другому розділі також проведений розвідувальний аналіз даних, в результаті

якого було побудовано статистичний звіт щодо показників датасету, виконана візуалізація кореляційних матриць Пірсона, Спірмена, Кендалла і Phik, за допомогою яких було визначено приблизну 93% кореляцію між цільовою змінною та фактором впливу.

3 СИСТЕМНИЙ АНАЛІЗ ТА ПЕРЕДБАЧЕННЯ ДОБОВОГО ГЕНЕРУВАННЯ ЕЛЕКТРОЕНЕРГІЇ ФОТОЕЛЕКТРОСТАНЦІЄЮ

3.1 Розділення вхідних даних на тренувальні, тестові та валідаційні

Розіб'ємо вхідні дані на два набори: тренувальний та тестовий, за допомогою команди «train_test_split». Обсяг тренувальних даних буде становить 60%, тоді як тестових - 40% (рис. 3.1).

```
# splitting base dataset into two: "training" and "test"
df = df_solar.copy()
X = df[['TOTAL_DAYS', 'TOTAL_TIME', 'SOLAR_RADIATION']]
y = df[['ENERGY']]
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=.4,random_state=0)
print(X_train.shape, X_test.shape)
```

(9222, 3) (6149, 3)

Рисунок 3.1 – Розподіл вхідних даних тренувальний та тестовий набори

В свою чергу тренувальний набір даних розділимо ще на два, виділив в ньому валідаційний набір, якій необхідний для вибору найкращої моделі [26]. Валідаційні дані вибираються у розмірі 20% від загальної кількості тренувальних даних (рис. 3.2).

```
# splitting "training" dataset into two: "validation" and "training"
X_train_all = X_train.copy()
y_train_all = y_train.copy()
X_train, X_valid, y_train, y_valid = train_test_split(X_train_all, y_train_all,
test_size=0.2, random_state=0)
```

Рисунок 3.2 – Розподіл тестових даних на валідаційний та тестові набори

На рисунку 3.3 показана статистика даних, які увійшли до тренувального набору.

```
# output "training" dataset
X_train.describe()
```

	TOTAL_DAYS	TOTAL_TIME	SOLAR_RADIATION
count	7377.000000	7377.000000	7377.000000
mean	177.786634	754.705165	374.269080
std	90.952943	210.728961	314.346153
min	1.000000	330.000000	10.000000
25%	106.000000	570.000000	89.000000
50%	177.000000	750.000000	280.000000
75%	246.000000	900.000000	640.000000
max	366.000000	1230.000000	1218.000000

Рисунок 3.3 – Дані які увійшли до тренувального набору

На рисунку 3.4 показана статистика даних, які увійшли до валідаційного набору.

```
# output "validation" dataset
X_valid.describe()
```

	TOTAL_DAYS	TOTAL_TIME	SOLAR_RADIATION
count	1845.000000	1845.000000	1845.000000
mean	180.657995	747.804878	357.159892
std	92.395762	211.585034	311.031467
min	1.000000	330.000000	11.000000
25%	107.000000	570.000000	79.000000
50%	178.000000	750.000000	255.000000
75%	252.000000	900.000000	608.000000
max	365.000000	1230.000000	1171.000000

Рисунок 3.4 – Дані які увійшли до валідаційного набору

На рисунку 3.5 показана статистика даних, які увійшли до тестового набору.

```
# output "test" dataset
X_test.describe()
```

	TOTAL_DAYS	TOTAL_TIME	SOLAR_RADIATION
count	6149.000000	6149.000000	6149.000000
mean	178.149943	753.898195	363.744999
std	92.391826	210.120972	311.358041
min	1.000000	330.000000	10.000000
25%	106.000000	570.000000	86.000000
50%	177.000000	750.000000	258.000000
75%	249.000000	930.000000	625.000000
max	366.000000	1230.000000	1245.000000

Рисунок 3.5 – Дані які увійшли до тестового набору

Маючи підготовлені набори даних можна приступати до подальшої побудови моделей передбачення об'ємів генерації активної енергії в зазначений проміжок часу, з урахуванням сонячної активності.

3.2 Побудова моделей передбачення генерування електроенергії та вибір найкращої з них

Для побудови моделей передбачення, будемо використовувати наступні методи: `DecisionTreeRegressor`, `RandomForestRegressor`, `LightGBMRegressor`. Для кожного методу створимо датафрейм, як показано на рисунку 3.6.

```

# Creation the dataframe with the resulting score of all models
result = pd.DataFrame({'model': [
                                'DecisionTreeRegressor',
                                'RandomForestRegressor',
                                'LightGBMRegressor'
                            ],
                       'train_score': 0,
                       'valid_score': 0})

result

```

	model	train_score	valid_score
0	DecisionTreeRegressor	0	0
1	RandomForestRegressor	0	0
2	LightGBMRegressor	0	0

Рисунок 3.6 – Створення датафрейму для різних моделей передбачення

Створим модель на основі метода `DecisionTreeRegressor`, виконаємо тренування та виведем її точність передбачення на тренувальному наборі (рис. 3.7), а потім на валідаційному (рис. 3.8).

```

from sklearn.tree import DecisionTreeRegressor

# Decision Tree Regressor
dtr = DecisionTreeRegressor()
dtr.fit(X_train,y_train['ENERGY'])

# Prediction for training dataset
y_pred_dtr = dtr.predict(X_train)

# Accuracy of model
r2_score_dtr = round(r2_score(y_train['ENERGY'], y_pred_dtr), 5)

# Save to result dataframe
result.loc[result['model'] == 'DecisionTreeRegressor', 'train_score'] = r2_score_dtr

print(f'DecisionTreeRegressor model, accuracy prediction for training dataset is: {r2_score_dtr}')

```

DecisionTreeRegressor model, accuracy prediction for training dataset is: 1.0

Рисунок 3.7 – Точність передбачення моделі DecisionTreeRegressor на тренувальному наборі

```

# Prediction for validation dataset
y_val_dtr = dtr.predict(X_valid)

# Accuracy of model
r2_score_dtr_valid = round(r2_score(y_valid['ENERGY'], y_val_dtr),5)

# Save to result dataframe
result.loc[result['model'] == 'DecisionTreeRegressor', 'valid_score'] = r2_score_dtr_valid

print(f'DecisionTreeRegressor model, accuracy prediction for validation dataset is: {r2_score_dtr_valid}')

```

DecisionTreeRegressor model, accuracy prediction for validation dataset is: 0.98703

Рисунок 3.8 – Точність передбачення моделі DecisionTreeRegressor на валідаційному наборі

Як видно з рисунків, отримана точність моделі на тренувальних даних - 1,0, а на валідаційних – 0,98703%.

Створим модель на основі метода RandomForestRegressor, виконаємо її тренування та виведем точність передбачення на тренувальному наборі (рис. 3.9), а потім на валідаційному (рис. 3.10).

```
from sklearn.ensemble import RandomForestRegressor

# Random Forest Regressor
rfr = RandomForestRegressor()

# Training model
rfr.fit(X_train,y_train['ENERGY'])

# Prediction for training dataset
y_pred_rfr = rfr.predict(X_train)

# Accuracy of model
r2_Score_rfr = round(r2_score(y_train['ENERGY'], y_pred_rfr), 5)

# Save to result dataframe
result.loc[result['model'] == 'RandomForestRegressor', 'train_score'] = r2_Score_rfr

print(f'RandomForestRegressor model, accuracy prediction for training dataset is: {r2_Score_rfr}')

RandomForestRegressor model, accuracy prediction for training dataset is: 0.99891
```

Рисунок 3.9 – Точність передбачення моделі RandomForestRegressor на тренувальному наборі

```
# Prediction for validation dataset
y_val_rfr = rfr.predict(X_valid)

# Accuracy of model
r2_score_rfr_valid = round(r2_score(y_valid['ENERGY'], y_val_rfr),5)

# Save to result dataframe
result.loc[result['model'] == 'RandomForestRegressor', 'valid_score'] = r2_score_rfr_valid

print(f'RandomForestRegressor model, accuracy prediction for validation dataset is: {r2_score_rfr_valid}')

RandomForestRegressor model, accuracy prediction for validation dataset is: 0.99273
```

Рисунок 3.10 – Точність передбачення моделі RandomForestRegressor на валідаційному наборі

Для цієї моделі бачимо точність на тренувальних даних – 0,99891, а на валідаційних – 0,99273%.

Створим модель на основі метода «LightGBMRegressor», виконаємо її тренування та виведем точність передбачення на тренувальному наборі (рис. 3.11), а потім на валідаційному (рис. 3.12).

```

import lightgbm as lgb
from lightgbm import LGBMClassifier

train_data = lgb.Dataset(X_train, label=y_train['ENERGY'])
params = {
    'num_leaves': 50,
    'learning_rate': 0.05,
    'metric': 'mae',
}

# Training model
model_lgb = lgb.train(params, train_data, num_boost_round=1000)

# Prediction for training dataset
y_pred_lgb = model_lgb.predict(X_train)

# Accuracy of model
r2_Score_lgb = round(r2_score(y_train['ENERGY'], y_pred_lgb), 5)

# Save to result dataframe
result.loc[result['model'] == 'LightGBMRegressor', 'train_score'] = r2_Score_lgb

print(f'LightGBMRegressor model, accuracy prediction for training dataset is: {r2_Score_lgb}')

```

[LightGBM] [Warning] Auto-choosing row-wise multi-threading, the overhead of testing was 0.003696 seconds. You can set `force_row_wise=true` to remove the overhead. And if memory is not enough, you can set `force_col_wise=true`.

[LightGBM] [Info] Total Bins 539

[LightGBM] [Info] Number of data points in the train set: 7377, number of used features: 3

[LightGBM] [Info] Start training from score 364.811577

LightGBMRegressor model, accuracy prediction for training dataset is: 0.99806

Рисунок 3.11 – Точність передбачення моделі LightGBMRegressor на тренувальному наборі

```

# Prediction for validation dataset
y_val_lgb = model_lgb.predict(X_valid)

# Accuracy of model
r2_score_lgb_valid = round(r2_score(y_valid['ENERGY'], y_val_lgb), 5)

# Save to result dataframe
result.loc[result['model'] == 'LightGBMRegressor', 'valid_score'] = r2_score_lgb_valid

print(f'LightGBMRegressor model, accuracy prediction for validation dataset is: {r2_score_lgb_valid}')

```

LightGBMRegressor model, accuracy prediction for validation dataset is: 0.99258

Рисунок 3.12 – Точність передбачення моделі LightGBMRegressor на валідаційному наборі

Для цієї моделі ми бачимо точність на тренувальних даних – 0,99806%, а на валідаційних – 0,99258%.

Після побудови моделей, необхідно обрати найкращу з них. Для цього, спочатку вибираються ті моделі, де різниця між тренувальними та валідаційними даними складає не більше ніж 10%. Після цього, серед них вибирається та, у якій найбільше значення параметру «valid_score» (рис. 3.13).

```
# Filtering models with big difference between training and validation datasets (dif > 10%)
result_best = result[(result['train_score'] - result['valid_score']).abs() < 10]
result_best.sort_values(by=['valid_score', 'train_score'], ascending=False)
```

	model	train_score	valid_score
1	RandomForestRegressor	0.99891	0.99273
2	LightGBMRegressor	0.99806	0.99258
0	DecisionTreeRegressor	1.00000	0.98703

```
# Choosing the best model
result_best = result.nlargest(1, 'valid_score')
result_best
```

	model	train_score	valid_score
1	RandomForestRegressor	0.99891	0.99273

Рисунок 3.13 – Визначення найкращої моделі

Отже, виходячи з результату, найкраща модель передбачення була побудована на основі метода RandomForestRegressor.

Виконаємо візуалізацію передбачень моделей на тренувальному,

валідаційному та тестовому наборах даних.

Фрагмент коду для побудови візуалізації моделей на тренувальному наборі даних наведено на рисунку 3.14.

```
# Model Predictions for training dataset

sampling_range_from = 100
sampling_range_to = 150

X_train = X_train[sampling_range_from:sampling_range_to]
y_train = y_train[sampling_range_from:sampling_range_to]

y_pred_dtr = y_pred_dtr[sampling_range_from:sampling_range_to]
y_pred_rfr = y_pred_rfr[sampling_range_from:sampling_range_to]
y_pred_lgb = y_pred_lgb[sampling_range_from:sampling_range_to]

x = np.arange(len(X_train))
plt.figure(figsize=(20,12))
plt.scatter(x, y_train['ENERGY'], label = "Target", color = 'g', marker = "x", s = 180)
# plt.scatter(x, y_pred_dtr, label = "DecisionTreeRegressor prediction", color = 'y', s = 30)
plt.scatter(x, y_pred_rfr, label = "RandomForestRegressor prediction", color = 'b', s = 45)
# plt.scatter(x, y_pred_lgb, label = "LightGBMRegressor prediction", color = 'brown', s = 30)
plt.plot(x, np.full(len(X_train), 1200), label = "Maximum allowable value", color = 'r')
plt.title('Model Predictions for training dataset')
plt.legend(loc='upper right')
plt.grid(True)
```

Рисунок 3.14 – Фрагмент коду для візуалізації моделей передбачень на тренувальному наборі

Візуалізація передбачення моделі на основі метода RandomForestRegressor для тренувального набору даних, наведено на рисунку 3.15.

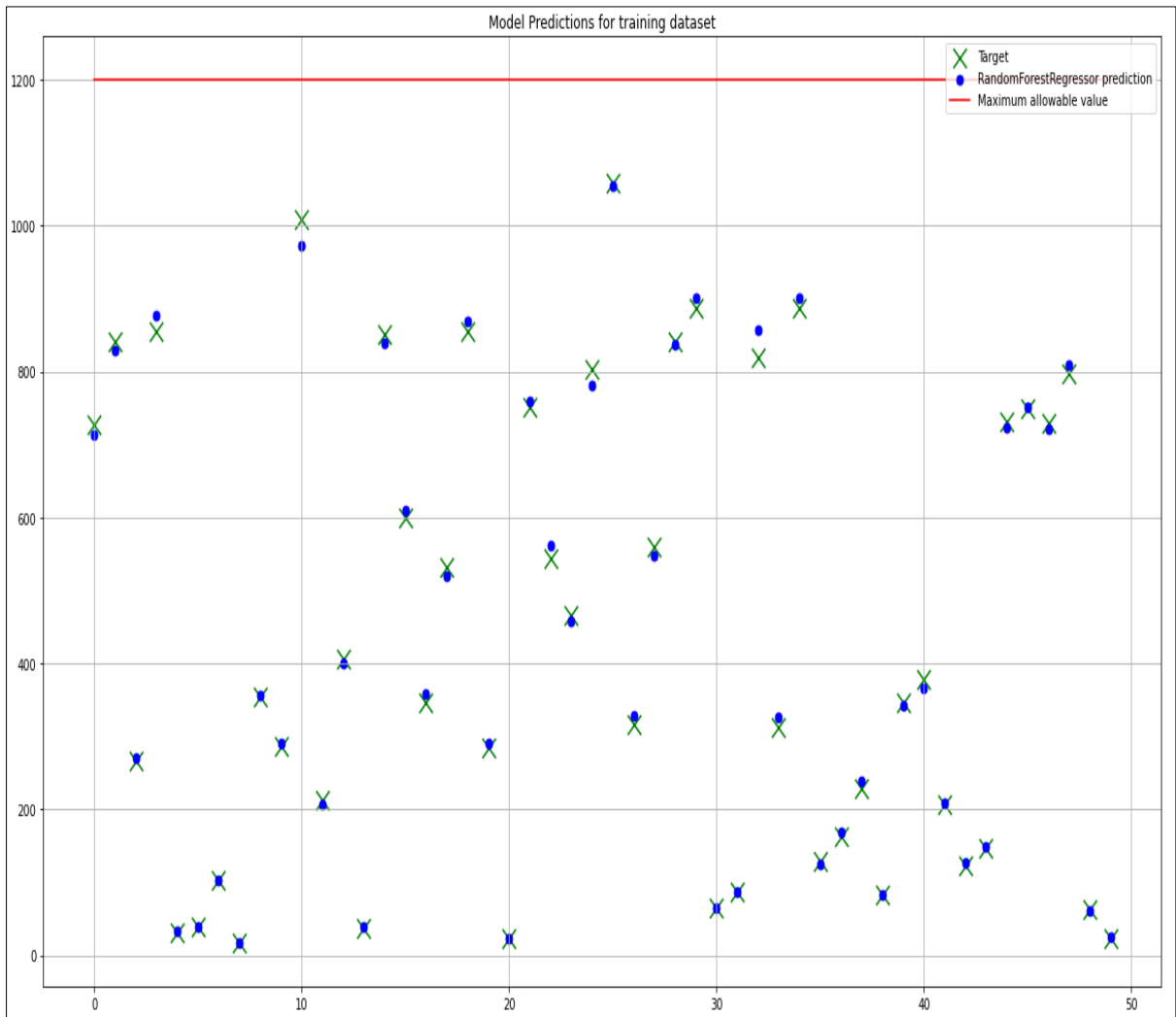


Рисунок 3.15 – Візуалізація передбачення моделі на основі метода RandomForestRegressor для тренувального набору даних

Фрагмент коду для побудови візуалізації моделей на валідаційному наборі даних наведено на рисунку 3.16.

```

# Model Predictions for validation dataset
sampling_range_from = 100
sampling_range_to = 150

X_valid = X_valid[sampling_range_from:sampling_range_to]
y_valid = y_valid[sampling_range_from:sampling_range_to]

y_val_dtr = y_val_dtr[sampling_range_from:sampling_range_to]
y_val_rfr = y_val_rfr[sampling_range_from:sampling_range_to]
y_val_lgb = y_val_lgb[sampling_range_from:sampling_range_to]

x = np.arange(len(X_valid))
plt.figure(figsize=(20,12))
plt.scatter(x, y_valid['ENERGY'], label = "Target", color = 'g', marker = "x", s = 180)
#plt.scatter(x, y_val_dtr, label = "DecisionTreeRegressor prediction", color = 'y', s = 30)
plt.scatter(x, y_val_rfr, label = "RandomForestRegressor prediction", color = 'b', s = 45)
#plt.scatter(x, y_val_lgb, label = "LightGBMRegressor prediction", color = 'brown', s = 30)
plt.plot(x, np.full(len(X_valid), 1200), label = "Maximum allowable value", color = 'r')
plt.title('Model Predictions for validation dataset')
plt.legend(loc='upper right')
plt.grid(True)

```

Рисунок 3.16 – Фрагмент коду для візуалізації моделей передбачень на валідаційному наборі

Візуалізація передбачення моделі на основі метода RandomForestRegressor для валідаційного набору даних, наведено на рисунку 3.17.

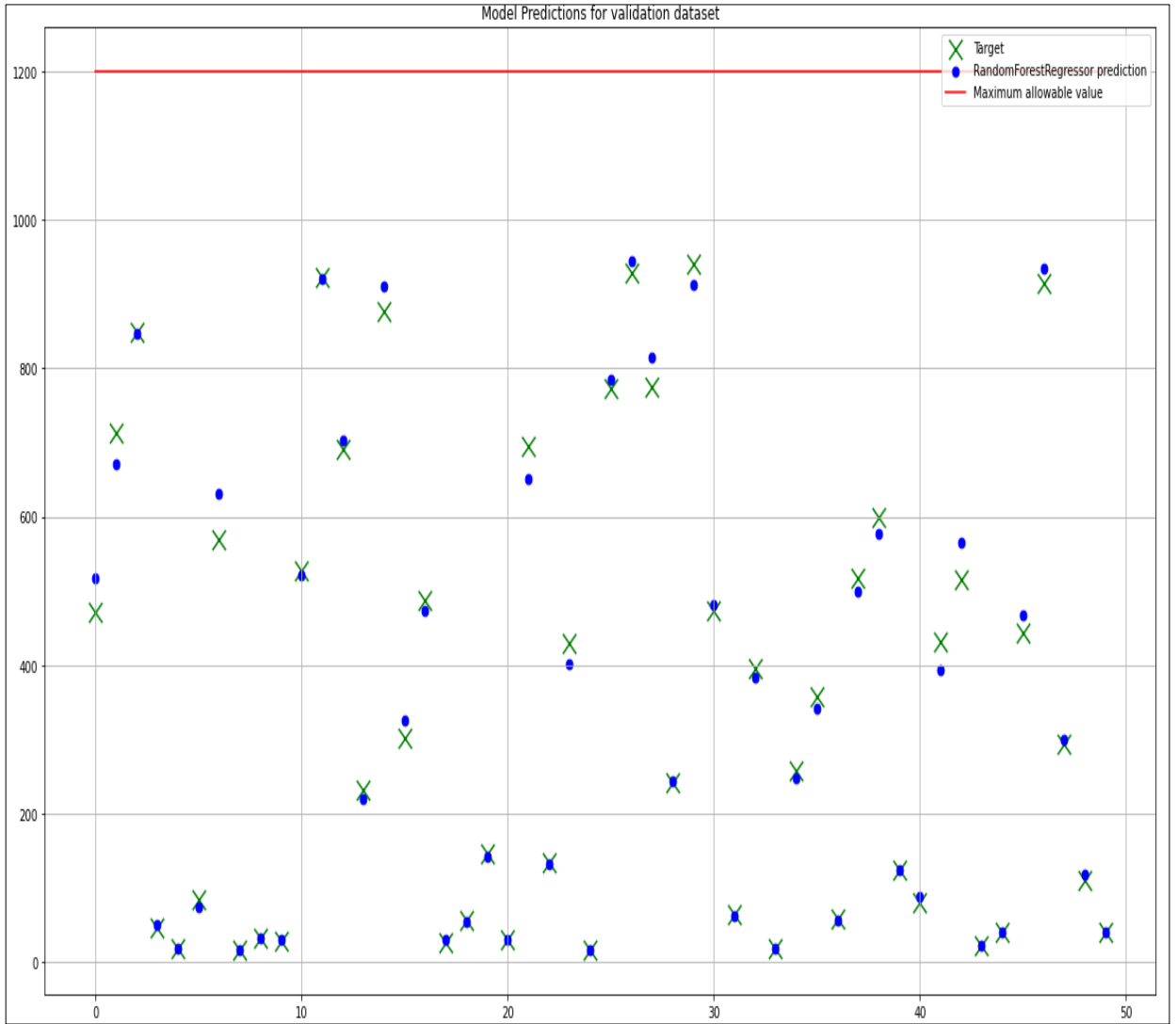


Рисунок 3.17 – Візуалізація передбачення моделі на основі метода RandomForestRegressor для валідаційного набору даних

Фрагмент коду для побудови візуалізації моделей на тестовому наборі даних наведено на рисунку 3.18.


```

# Prediction targets for test dataset
sampling_range_from = 100
sampling_range_to = 150

X_test = X_test[sampling_range_from:sampling_range_to]
y_test = y_test[sampling_range_from:sampling_range_to]

y_test_dtr = dtr.predict(X_test)
y_test_rfr = rfr.predict(X_test)
y_test_lgb = model_lgb.predict(X_test)

# Model Predictions for test dataset
x = np.arange(len(X_test))
plt.figure(figsize=(20,12))
plt.scatter(x, y_test['ENERGY'], label = "Target", color = 'g', marker = "x", s = 180)
# plt.scatter(x, y_test_dtr, label = "DecisionTreeRegressor prediction", color = 'y', s = 30)
plt.scatter(x, y_test_rfr, label = "RandomForestRegressor prediction", color = 'b', s = 45)
# plt.scatter(x, y_test_lgb, label = "LightGBMRegressor prediction", color = 'brown', s = 30)
plt.plot(x, np.full(len(X_test), 1200), label = "Maximum allowable value", color = 'r')
plt.title('Prediction for the test data')
plt.legend(loc='upper right')
plt.grid(True)

```

Рисунок 3.18 – Фрагмент коду для візуалізації моделей передбачень на тестовому наборі

Візуалізація передбачення моделі на основі метода RandomForestRegressor для тестового набору даних, наведено на рисунку 3.19.

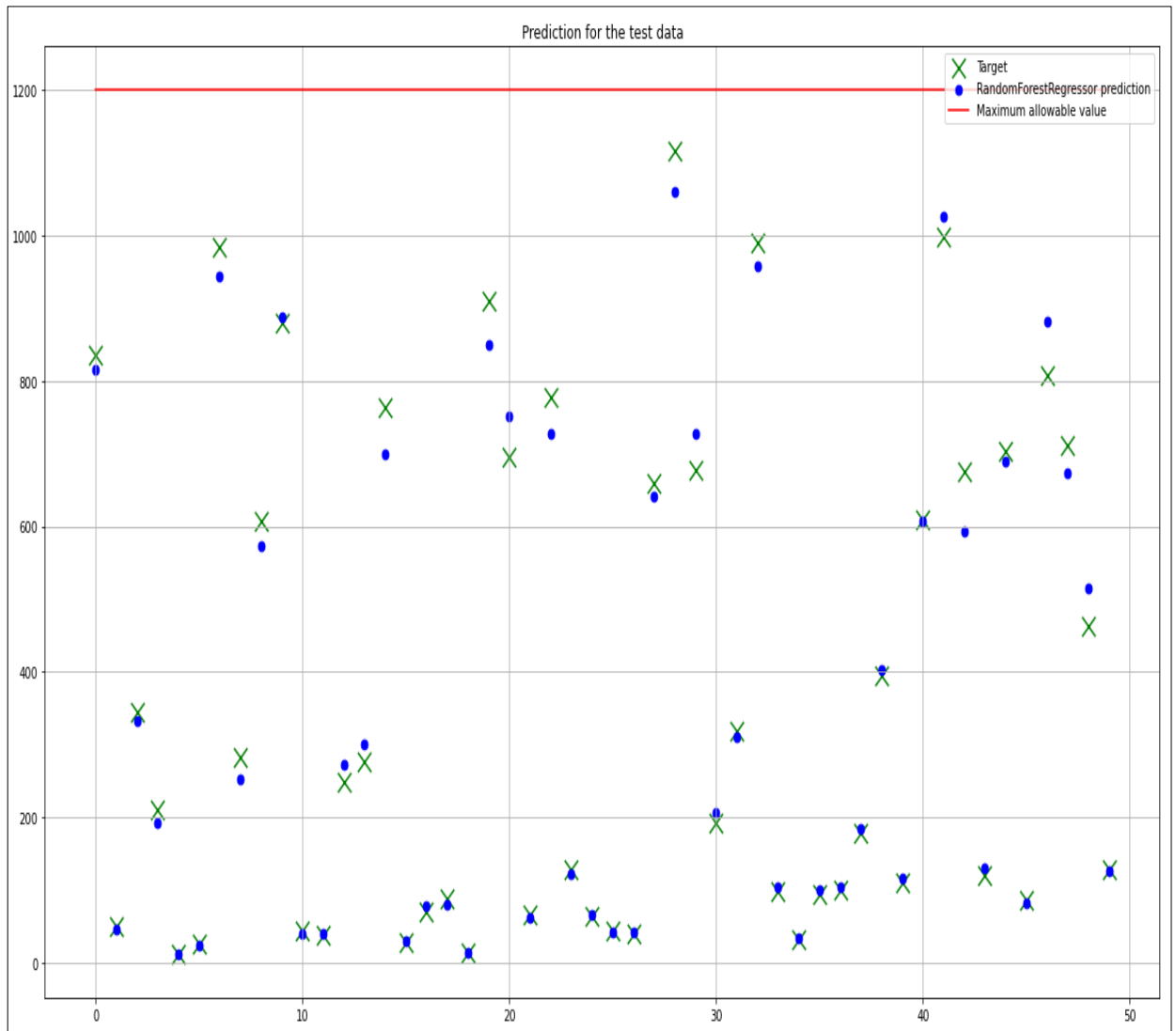


Рисунок 3.19 – Візуалізація передбачення моделі на основі метода RandomForestRegressor для тестового набору даних

3.3 Висновки

В третьому розділі магістерської кваліфікаційної роботи побудовано декілька моделей передбачення на основі методів: DecisionTreeRegressor, RandomForestRegressor, LightGBMRegressor. Проведено їх навчання та перевірка точності їх передбачення, що супроводжувалося відповідними візуальними діаграмами. По результатах дослідження обрано найкращу модель, якою виявилась модель, що побудована на основі метода RandomForestRegressor з точністю передбачення 0.99277. Така модель здатна

з великою точністю передбачати добову генерацію електроенергії, що буде мати вагомий ефект для фотоелектричної станції.

4 ЕКОНОМІЧНА ЧАСТИНА

4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки

Процес виробництва електроенергії на фотоелектричних станціях є повністю екологічним і не створює забруднюючих елементів для навколишнього середовища, а також є одним з найефективніших відновлюваних джерел енергії, які існують на даний момент. Завдяки перевагам сонячної енергії в порівнянні з енергією, виробленою з викопного палива або невідновлюваних джерел, сонячні електростанції є ключовим інструментом для розробки нової моделі довгострокового сталого виробництва, яка є повністю екологічною. Вважається, що потенціал відновлюваних джерел енергії здатен забезпечити достатнє та надійне електропостачання у всьому світі до 2050 року. Це доводять результати моделювання Технологічного університету Лапенранта (Фінляндія). Така перспектива спонукає державних та приватних інвесторів мобілізувати мільярди доларів на вдосконалення технологій відновлюваної енергетики.

Але сонячна енергетика є суттєво залежною від погодних факторів, а отже є нестабільною, що може призводити до розбалансування енергосистеми. В Україні електрогенеруючі компанії обкладаються досить великими штрафами за недотримання балансу з вироблення електроенергії. В зв'язку з цим, на сьогоднішній день проблеми прогнозування або передбачення генерації фотоелектричними станціями є актуальним та важливими. Велика увага приділяється аудиту технологій та розробок у сфері сонячної енергії, зокрема генерації.

Критерії оцінювання наукового-технічного рівня та комерційного потенціалу розробки наведено в таблиці 4.1 [42].

Таблиця 4.1 – Критерії оцінювання наукового-технічного рівня та комерційного потенціалу розробки та бальна оцінка

Бали (за п'ятибальною шкалою)					
	0	1	2	3	4
<i>l</i>	2	3	4	5	6
<i>Технічна здійсненність концепції</i>					
1	Достовірність Концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена на практиці	Перевірено роботоздатність продукту в реальних умовах
<i>Ринкові переваги (недоліки)</i>					
2	Багато аналогів на малому ринку	Мало аналогів на малому ринку	Кілька аналогів на великому ринку	Один аналог на великому ринку	Продукт не має аналогів на великому ринку
3	Ціна продукту значно вища за ціни аналогів	Ціна продукту дещо вища за ціни аналогів	Ціна продукту приблизно дорівнює цінам аналогів	Ціна продукту дещо нижча за ціни аналогів	Ціна продукту значно нижча за ціни аналогів
4	Технічні та споживчі властивості продукту значно гірші, ніж в аналогів	Технічні та споживчі властивості продукту трохи гірші, ніж в аналогів	Технічні та споживчі властивості продукту нарівні аналогів	Технічні та споживчі властивості продукту трохи кращі, ніж в аналогів	Технічні та споживчі властивості продукту значно кращі, ніж в аналогів
5	Експлуатаційні витрати значно вищі, ніж в аналогів	Експлуатаційні витрати дещо вищі, ніж в аналогів	Експлуатаційні витрати на рівні експлуатаційних витрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значно нижчі, ніж в аналогів
<i>Ринкові перспективи</i>					
6	Ринок малий і не має позитивної динаміки	Ринок малий, але має позитивну динаміку	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою
7	Активна конкуренція великих компаній на ринку	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкуренція немає

Продовження таблиці 4.1

Бали (за п'ятибальною шкалою)					
	0	1	2	3	4
1	2	3	4	5	6
<i>Практична здійсненність</i>					
8	Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї	Необхідно наймати фахівців або витратити значні кошти та час на навчання наявних фахівців	Необхідне незначне навчання фахівців та збільшення їх штату	Необхідне незначне навчання фахівців	Є фахівці з питань як з технічної, так і з комерційної реалізації ідеї
9	Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні	Потрібні незначні фінансові ресурси. Джерела фінансування відсутні	Потрібні значні фінансові ресурси. Джерела фінансування є	Потрібні незначні фінансові ресурси. Джерела фінансування є	Не потребує додаткового фінансування
10	Необхідна розробка нових матеріалів	Потрібні матеріали, що використовуються у військово-промисловому комплексі	Потрібні дорогі матеріали	Потрібні досяжні та дешеві матеріали	Всі матеріали для реалізації ідеї відомі та давно використовуються у виробництві
11	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років	Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більший 5-ти років	Термін реалізації ідеї менший 3-х років. Термін окупності інвестицій від 3-х до 5-ти років	Термін реалізації ідеї менший 3-х років. Термін окупності інвестицій менший 3-х років
12	Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію продукту	Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що потребує значних коштів та часу	Процедура отримання дозвільних документів для виробництва та реалізації продукту потребує незначних коштів та часу	Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту	Відсутні будь-які регламентні обмеження на виробництво та реалізацію продукту

Результати оцінювання комерційного потенціалу розробки наведено в таблиці 4.2

Таблиця 4.2 Результат оцінювання комерційного потенціалу розробки

Критерії	Експерт		
	1	2	3
	Бали:		
1. Технічна здійсненність концепції	4	3	3
2. Ринкові переваги (наявність аналогів)	4	4	3
3. Ринкові переваги (ціна продукту)	4	3	3
4. Ринкові переваги (технічні властивості)	3	4	3
5. Ринкові переваги (експлуатаційні витрати)	3	2	3
6. Ринкові перспективи (розмір ринку)	3	4	4
7. Ринкові перспективи (конкуренція)	3	4	3
8. Практична здійсненність (наявність фахівців)	3	3	2
9. Практична здійсненність (наявність фінансів)	3	3	3
10. Практична здійсненність (необхідність нових матеріалів)	2	3	3
11. Практична здійсненність (термін реалізації)	3	2	3
12. Практична здійсненність (розробка документів)	4	3	4
Сума балів	СБ1=39	СБ2=38	СБ3=37
Середньоарифметична сума балів СБс	$СБ_c = \frac{\sum_{i=1}^3 СБ_i}{3} = \frac{114}{3} = 38$		

За даними таблиці 4.2 можна зробити висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки. Для цього доцільно скористатись рекомендаціями, наведеними в таблиці 4.3 [42].

Таблиця 4.3 Науково-технічні рівні та комерційні потенціали розробки

Середньоарифметична сума балів СБ,розрахована на основі висновків експертів	Науково-технічний рівень та комерційний потенціал розробки
41...48	Високий
31...40	Вищий середнього
21...30	Середній
11...20	Нижчий середнього
0...10	Низький

З таблиці видно, що науково-технічний рівень та комерційний потенціал розробки є вищим середнього. Це досягається за рахунок обов'язків оперативного персоналу фотоелектростанції надавати погодинне передбачення генерації електроенергії на наступну добу по доведеному прогнозу хмарності. У випадку відхилення від заданого графіка більш ніж на 10%, підприємство, що експлуатує електростанцію, має сплачувати штрафи у розмірі 100% вартості електроенергії. Точне передбачення генерації стає вельми важливим.

Вищий ніж середній рівень комерційного потенціалу розробленої інформаційної технології досягається завдяки використанню методів машинного навчання. Ці методи дозволяють підвищити точність передбачення порівняно з аналогічними технологіями. [42].

4.2 Розрахунок витрат на проведення науково-технічної роботи

4.2.1 Витрати на оплату праці

Основна заробітна плата дослідників.

Витрати на основну заробітну плату дослідників (Z_0) розраховують відповідно до посадових окладів працівників, за формулою [42]:

$$Z_0 = \sum_{i=1}^k \frac{M_{ni} \cdot t_i}{T_p}, \quad (4.1)$$

де k – кількість посад дослідників, залучених до процесу досліджень;

M_{ni} – місячний посадовий оклад конкретного дослідника, грн;

t_i – кількість днів роботи конкретного дослідника, дн.;

T_p – середня кількість робочих днів в місяці, $T_p=21\dots23$ дні.

Проведені розрахунки зводимо до таблиці 4.4.

Таблиця 4.4 – Витрати на заробітну плату дослідників

Найменування посади	Місячний посадовий оклад, грн	Оплата за робочий день, грн	Кількість днів роботи	Витрати на заробітну плату, грн
Керівник проекту	33000	1500	50	75000
Програміст	27500	1250	50	62500
Всього				137500

Основна заробітна плата робітників

Витрати на основну заробітну плату робітників (Z_p) за відповідними найменуваннями робіт на тему «Аналіз та передбачення добового генерування електроенергії фотоелектростанцією» розраховується за формулою:

$$Z_p = \sum_{i=1}^n C_i * t_i, \quad (4.2)$$

де C_i – погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

t_i – час роботи робітника на виконання певної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду C_i можна визначити за формулою:

$$C_i = \frac{M_M \cdot K_i \cdot K_C}{T_p \cdot t_{3M}}, \quad (4.3)$$

де M_M – розмір прожиткового мінімуму працездатної особи або мінімальної місячної заробітної плати (залежно від діючого законодавства), приймемо значення 6700 грн;

K_i – коефіцієнт міжкваліфікаційного співвідношення для встановлення тарифної ставки робітнику відповідного розряду [42];

K_C – мінімальний коефіцієнт співвідношень місячних тарифних ставок робітників першого розряду з нормальними умовами праці виробничих об'єднань і підприємств до законодавчо встановленого розміру мінімальної заробітної плати [42];

T_p – середня кількість робочих днів в місяці, приблизно $T_p = 21 \dots 23$ дні;

t_{3M} – тривалість зміни, год.

Обчислимо погодинну тарифну ставку робітника.

$$C_I = \frac{6700 \cdot 1,1 \cdot 1,65}{21 \cdot 8} = 72,38 \text{ (грн)}.$$

Визначимо витрати на основну заробітну плату робітників

$$Z_p = 72,38 \cdot 8,2 = 593,52 \text{ (грн)}.$$

З урахуванням витрат на підготовку робочого місця, інсталяцію програмного забезпечення та компіляцію програмних частин, загальна величини витрат на основну заробітну плату робітників становить 2094,83 грн. Зазначені витрати вказані в таблиці 4.5.

Таблиця 4.5 – Величина витрат на основну заробітну плату робітників

Найменування робіт	Тривалість роботи, год	Розряд роботи	Тарифний коефіцієнт	Погодинна тарифна ставка, грн	Величина оплати на робітника, грн
Підготовка робочого місця	8,2	2	1,1	72,38	593,52
Інсталяція програмного забезпечення	6,7	3	1,35	88,83	595,161
Компіляція програмних частин	8,1	5	1,7	111,87	906,147
Всього					2094,83

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 ... 12% від суми основної заробітної плати дослідників та робітників за формулою:

$$Z_{\text{дод}} = (Z_o + Z_p) \cdot \frac{N_{\text{дод}}}{100\%}, \quad (4.4)$$

де $N_{\text{дод}}$ – норма нарахування додаткової заробітної плати.

$$Z_{\text{дод}} = (137500 + 2094,83) \cdot \frac{10}{100\%} = 13959,48 \text{ (грн)}.$$

4.2.2 Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуємо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою:

$$З_{\text{н}} = (З_{\text{о}} + З_{\text{р}} + З_{\text{дод}}) \cdot \frac{Н_{\text{зп}}}{100\%}, \quad (4.5)$$

де $Н_{\text{зп}}$ – норма нарахування на заробітну плату.

$$З_{\text{н}} = (137500 + 2094,83 + 13959,48) \cdot \frac{22}{100\%} = 33781,95 \text{ (грн)}.$$

4.2.3 Сировина та матеріали

Витрати на матеріали (M) у вартісному вираженні розраховуємо окремо для кожного виду матеріалів за формулою:

$$M = \sum_{j=1}^n H_j \cdot Ц_j - \sum_{j=1}^n B_j \cdot Ц_j, \quad (4.6)$$

де H_j – норма витрат матеріалу j -го найменування, кг;

n – кількість видів матеріалів;

$Ц_j$ – вартість матеріалу j -го найменування, грн/кг;

K_j – коефіцієнт транспортних витрат, ($K_j = 1,1 \dots 1,15$);

B_j – маса відходів j -го найменування, кг;

$Ц_{\text{в}j}$ – вартість відходів j -го найменування, грн/кг.

Проведені розрахунки зведемо в таблицю 4.6.

Таблиця 4.6 – Витрати на матеріали

Найменування матеріалу, марка, тип, сорт	Ціна за 1 кг, грн	Норма витрат, кг	Величина відходів, кг	Ціна відходів, грн/кг	Вартість витраченого матеріалу, грн
Flash-пам'ять 256GB	540,0	1,0	-	-	621,00
Офісний папір А4 500	210,0	3,0	-	-	724,50
Картридж для принтера	950,0	1,0	-	-	1092,50
Набір канцелярський офісний	310,0	2,0	-	-	713
Органайзер офісний	480,0	3,0	-	-	1656
Папір для записів А5 250	95,0	3,0	-	-	327,75
Всього					5134,75

4.2.4 Розрахунок витрат на комплектуючі

Витрати на комплектуючі відсутні.

4.2.5 Спецустаткування для наукових (експериментальних) робіт

Витрати на спецустаткування для наукових (експериментальних) робіт відсутні.

4.2.6 Програмне забезпечення для наукових (експериментальних) робіт

Витрати на програмне забезпечення для наукових (експериментальних) робіт відсутні.

4.2.7 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою:

$$A_{\text{обл}} = \frac{Ц_{\text{б}}}{T_{\text{в}}} \cdot \frac{t_{\text{вик}}}{12}, \quad (4.7)$$

де $Ц_{\text{б}}$ – балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{\text{вик}}$ – термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

$T_{\text{в}}$ – строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

Розрахуємо амортизаційні відрахування. Проведені розрахунки зведено в таблиці 4.7.

$$A_{\text{обл}} = \frac{26300 \cdot 3}{2 \cdot 12} = 3287,50 \text{ (грн)}.$$

Таблиця 4.7 – Амортизація відрахування по кожному виду обладнання

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Персональний комп'ютер проведення	26300	2	3	3287,50

розробки та моделювання				
-------------------------	--	--	--	--

Продовження таблиці 4.7.

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Робоче місце інженера розробника	9500	5	3	475
Офісне приміщення	1300000	20	3	16250
ОС Windows 10 PRO	7000	2	3	875
Прикладний пакет Microsoft Office 2021	12000	2	3	1500
Всього				22387,50

4.2.8 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію (B_e) розраховуємо за формулою:

$$B_e = \sum_{i=1}^n \frac{W_{yi} \cdot t_i \cdot C_e \cdot K_{впi}}{\eta_i}, \quad (4.8)$$

де W_{yi} – встановлена потужність обладнання на визначеному етапі розробки, кВт;

t_i – тривалість роботи обладнання на етапі дослідження, год;

C_e – вартість 1 кВт*год електроенергії, грн; прийmemo $C_e = 7,50$ грн;

$K_{впi}$ – коефіцієнт, що враховує використання потужності, $K_{впi} < 1$;

η_i – коефіцієнт корисної дії обладнання, $\eta_i < 1$.

$$B_{\epsilon} = 0,32 \cdot 360,0 \cdot 7,50 \cdot 0,96 / 0,97 = 855,09 \text{ (грн)}.$$

Проведені розрахунки зведемо до таблиці 4.8.

Таблиця 4.8 – Витрати на електроенергію

Найменування обладнання	Встановлена потужність, кВт	Тривалість роботи, год	Сума, грн
Персональний комп'ютер проведення розробки та моделювання	0,32	360,0	855,09
Робоче місце інженера	0,12	310,0	276,12
Всього			1131,21

4.2.9 Службові відрядження

Витрати за статтею «Службові відрядження» розраховуємо як 20...25% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{\text{св}} = (Z_0 + Z_p) \cdot \frac{N_{\text{св}}}{100\%}, \quad (4.9)$$

де $N_{\text{св}}$ – норма нарахування за статтею «Службові відрядження», прийmemo $N_{\text{св}} = 25\%$.

$$B_{\text{св}} = (150000 + 2094,83) \cdot \frac{25}{100\%} = 38023,7 \text{ (грн)}.$$

4.2.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації

Витрати на роботи, які виконують сторонні підприємства, установи і організації відсутні.

4.2.11 Інші витрати

Витрати за статтею «Інші витрати» розраховуємо як 50...100% від суми основної заробітної плати дослідників та робітників за формулою:

$$I_{\text{в}} = (Z_{\text{о}} + Z_{\text{р}}) \cdot \frac{H_{\text{ів}}}{100\%}, \quad (4.10)$$

де $H_{\text{ів}}$ – норма нарахування за статтею «Інші витрати», прийmemo $H_{\text{ів}} = 50\%$.

$$I_{\text{в}} = (137500 + 2094,83) \cdot \frac{50}{100\%} = 69797,42 \text{ (грн)}.$$

4.2.12 Накладні (загальновиробничі) витрати

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100...150% від суми основної заробітної плати дослідників та робітників за формулою:

$$V_{\text{нзв}} = (Z_{\text{о}} + Z_{\text{р}}) \cdot \frac{H_{\text{нзв}}}{100\%}, \quad (4.11)$$

де $H_{\text{нзв}}$ – норма нарахування за статтею «Накладні (загальновиробничі) витрати», прийmemo $H_{\text{нзв}} = 100\%$.

$$V_{\text{нзв}} = (137500 + 2094,83) \cdot 100 / 100\% = 139594,83 \text{ (грн)}.$$

Загальні витрати на проведення науково-технічної роботи на тему «Аналіз та прогнозування добового генерування електроенергії» розраховуємо як суму всіх попередніх статей витрат за формулою:

$$V_{\text{заг}} = Z_0 + Z_p + Z_{\text{дод}} + Z_n + M + K_v + V_{\text{спец}} + V_{\text{прг}} + A_{\text{обл}} + V_e + V_{\text{св}} + V_{\text{сп}} + I_v + V_{\text{нзв}} \quad (4.12)$$

$$V_{\text{заг}} = 137500 + 2094,83 + 13959,48 + 33781,95 + 5134,75 + 0 + 0 + 0 + 3287,50 + 1131,21 + 38023,7 + 0 + 69797,42 + 139594,83 = 444305,67 \text{ (грн).}$$

Загальні витрати на завершення науково-технічної роботи та оформлення її результатів розраховується за формулою:

$$ЗВ = \frac{V_{\text{заг}}}{\eta}, \quad (4.13)$$

де η - коефіцієнт, який характеризує етап (стадію) виконання науково-технічної роботи, прийmemo $\eta=0,90$.

$$ЗВ = \frac{444305,67}{0,90} = 493672,97 \text{ (грн).}$$

4.3 Розрахунок економічної ефективності науково-технічної розробки за її можливої комерціалізації потенційним інвестором

У ринкових умовах головною перевагою, яку потенційний інвестор може отримати від впровадження науково-технічної розробки, є зростання чистого прибутку. Цей ріст чистого прибутку призведе до додаткових

фінансових ресурсів для інвестора, покращення фінансових результатів його діяльності, підвищення конкурентоспроможності та можливе позитивне вплив на його рішення щодо комерціалізації розробки.

Для оцінки потенційного зростання чистого прибутку від впровадження науково-технічної розробки потрібно:

а) вказати, з якого часу можуть бути впроваджені результати науково-технічної розробки;

б) зазначити, протягом скількох років після впровадження цієї науково-технічної розробки очікуються основні позитивні результати для потенційного інвестора (наприклад, протягом 4-х років після її впровадження);

в) кількісно оцінити величину існуючого та майбутнього попиту на цю або аналогічні чи подібні науково-технічні розробки та назвати основних суб'єктів (зацікавлених осіб) цього попиту;

г) визначити ціну реалізації на ринку науково-технічних розробок з аналогічними чи подібними функціями.

Так, при оцінці економічної ефективності важливо враховувати концепцію зміни вартості грошей у часі. Це пов'язано з тим, що вартість грошей змінюється з плином часу через інфляцію або можливу отриману вартість в майбутньому. Врахування часової цінності грошей дозволяє коректно оцінити сьгоднішні інвестиції та їхні майбутні вигоди або втрати.

При оцінюванні ефективності інноваційних проектів передбачається розрахунок таких важливих показників:

- абсолютного економічного ефекту (чистого дисконтованого доходу);
- внутрішньої економічної дохідності (внутрішньої норми дохідності);
- терміну окупності (дисконтованого терміну окупності).

Проведення науково-технічних досліджень та оцінка їхньої економічної ефективності для можливої комерціалізації потенційним інвестором може включати аналіз різних напрямів, зокрема, фокусуючись на розробці або значному удосконаленні програмного забезпечення для масового споживача.

При розробці чи удосконаленні програмного засобу головною метою є створення продукту, який задовольнить потреби широкого ринку. Потенційний інвестор може розглядати цей вид розробки як інвестицію з високим потенціалом прибутку, оскільки успішне впровадження такого продукту може мати значний попит і приносити великий дохід на ринку.

В цьому випадку майбутній економічний ефект буде формуватися на основі таких даних [42]:

$$\Delta\Pi_i = (\pm\Delta\Pi_0 \cdot N + \Pi_0 \cdot \Delta N_i)_i \cdot \lambda \cdot \rho \cdot \left(1 - \frac{\vartheta}{100}\right), \quad (4.15)$$

де $\pm\Delta\Pi_0$ – зміна основного якісного показника від впровадження результатів науково-технічної розробки в аналізованому році. Зазвичай, таким показником може бути зміна ціни реалізації одиниці нової розробки в аналізованому році (відносно року до впровадження цієї розробки); $\pm\Delta\Pi_0$ може мати як додатне, так і від’ємне значення (від’ємне – при зниженні ціни відносно року до впровадження цієї розробки, додатне – при зростанні ціни). Прийmemo зростання на 200 грн;

N – основний кількісний показник, який визначає величину попиту на аналогічні чи подібні розробки у році до впровадження результатів нової науково-технічної розробки. Кількість споживачів прийmemo 27000 осіб;

Π_0 – основний якісний показник, який визначає ціну реалізації нової науково-технічної розробки в аналізованому році, $\Pi_0 = \Pi_6 \pm \Delta\Pi_0$;

Π_6 – основний якісний показник, який визначає ціну реалізації існуючої (базової) науково-технічної розробки у році до впровадження результатів. Прийmemo 2200,00 грн;

ΔN – зміна основного кількісного показника від впровадження результатів науково-технічної розробки в аналізованому році. Зазвичай таким показником може бути зростання попиту на науково-технічну розробку в аналізованому році (відносно року до впровадження цієї розробки);

λ – коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2023 році ставка податку на додану вартість становить 18%, а коефіцієнт $\lambda = 0,8333$;

ρ – коефіцієнт, який враховує рентабельність інноваційного продукту (послуги). Рекомендується брати $\rho = 0,2 \dots 0,5$. Прийmemo $\rho = 0,4$;

ϑ – ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2023 році $\vartheta = 18\%$

Розрахуємо збільшення чистого прибутку в перші чотири роки.

Збільшення чистого прибутку 1-го року: $\Delta\Pi_1 = (200 \cdot 27000 + 2400 \cdot 3000) \cdot 0,83 \cdot 0,4 \cdot \left(1 - \frac{18\%}{100\%}\right) = 3430224,00$ (грн).

Збільшення чистого прибутку 2-го року: $\Delta\Pi_2 = (200 \cdot 27000 + 2400 \cdot 4500) \cdot 0,83 \cdot 0,4 \cdot \left(1 - \frac{18\%}{100\%}\right) = 4410288,00$ (грн).

Збільшення чистого прибутку 3-го року: $\Delta\Pi_3 = (200 \cdot 27000 + 2400 \cdot 7200) \cdot 0,83 \cdot 0,4 \cdot \left(1 - \frac{18\%}{100\%}\right) = 6174403,20$ (грн).

Збільшення чистого прибутку 4-го року: $\Delta\Pi_4 = (200 \cdot 27000 + 2400 \cdot 10300) \cdot 0,83 \cdot 0,4 \cdot \left(1 - \frac{18\%}{100\%}\right) = 8199868,80$ (грн).

Далі розраховуємо приведену вартість збільшення всіх чистих прибутків ПП, що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ПП = \sum_{i=1}^T \frac{\Delta\Pi_i}{(1+\tau)^t}, \quad (4.16)$$

де $\Delta\Pi_i$ – збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

T – період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки,

роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні, $\tau = 0,05 \dots 0,15$. Прийmemo $\tau = 0,15$;

t – період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$$\begin{aligned} \text{ПП} &= \frac{3430224,00}{(1+0,15)^1} + \frac{4410288,00}{(1+0,15)^2} + \frac{6174403,20}{(1+0,15)^3} + \frac{8199868,80}{(1+0,15)^4} = 2982803,48 + \\ &3334811,34 + 4059770,33 + 4688301,6 = 15065686,8 \text{ (грн)}. \end{aligned}$$

Далі розраховують величину початкових інвестицій PV , які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки. Для цього можна використати формулу:

$$PV = k_{\text{інв}} \cdot 3B, \quad (4.17)$$

де $k_{\text{інв}}$ – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію. Це можуть бути витрати на підготовку приміщень, розробку технологій, навчання персоналу, маркетингові заходи тощо; зазвичай $k_{\text{інв}} = 2 \dots 5$, але може бути і більшим. Прийmemo $k_{\text{інв}} = 2,1$;

$3B$ – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, складають 530496,55 грн.

$$PV = k_{\text{інв}} \cdot 3B = 2,1 \cdot 530496,55 = 1114042,76 \text{ (грн)}.$$

Тоді абсолютний економічний ефект $E_{\text{абс}}$ або чистий приведений дохід (NPV, NetPresentValue) для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{абс} = ПП - PV, \quad (4.18)$$

де ПП – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 15065686,8 грн;

$$E_{абс} = 15065686,8 - 1114042,76 = 13951644,04 \text{ (грн)}.$$

Внутрішня економічна дохідність інвестицій E_B , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки, розраховується за формулою:

$$E_B = \sqrt[T_{ж}]{1 + \frac{E_{абс}}{PV}} - 1, \quad (4.19)$$

де $E_{абс}$ – абсолютний економічний ефект вкладених інвестицій, 13951644,04 грн;

PV – теперішня вартість початкових інвестицій, 1114042,76 грн;

$T_{ж}$ – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримання позитивних результатів від її впровадження, 4 роки.

$$\begin{aligned} E_B &= \sqrt[4]{1 + \frac{E_{абс}}{PV}} - 1 = \left(1 + \frac{13951644,04}{1114042,76}\right)^{\frac{1}{4}} - 1 = (12,52)^{\frac{1}{4}} - 1 = 1,88 - 1 \\ &= 0,88. \end{aligned}$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій $\tau_{мін}$ визначається за формулою:

$$\tau_{\text{мін}} = d + f, \quad (4.20)$$

де d – середньозважена ставка за депозитними операціями в комерційних банках; в 2023 році в Україні $d = 0,16$;

f – показник, що характеризує ризикованість вкладення інвестицій; зазвичай величина $f = 0,05 \dots 0,5$, але може бути і значно вищою. Прийmemo 0,23.

$\tau_{\text{мін}} = d + f = 0,16 + 0,23 = 0,39 < 0,88$, що свідчить про те, що внутрішня економічна дохідність інвестицій E_v , вища мінімальної внутрішньої дохідності. Тобто, потенційних інвестор може бути зацікавлений у фінансуванні науково-технічної розробки за темою «Аналіз та передбачення добового генерування електроенергії фотоелектростанцією» та виведенні її на ринок, тобто в її комерціалізації.

Період окупності інвестицій $T_{\text{ок}}$ (DPP, DiscountedPaybackPeriod), які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{\text{ок}} = \frac{1}{E_v}, \quad (4.21)$$

де E_v – внутрішня економічна дохідність вкладених інвестицій.

$$T_{\text{ок}} = \frac{1}{0,88} = 1,34 \text{ (р)}.$$

Як бачимо термін окупності становить менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок [42].

4.4 Висновки

Проведені економічні розрахунки підтверджують суттєвий комерційний потенціал розробки, яка фокусується на аналізі та передбаченні добового виробництва електроенергії з урахуванням сонячної активності. Оціночний показник комерційного потенціалу становить 38 балів, що свідчить про значні можливості для успішної комерціалізації цього науково-технічного рішення.

Варто зазначити, що відшкодування витрат на розробку відбудеться за 1,34 року. Це свідчить про високу комерційну привабливість розробки та швидке повернення грошей інвестора. Такий короткий термін має стимулювати потенційних інвесторів забезпечувати фінансування даної розробки та виведення її на ринок з подальшим отриманням прибутку.

Виходячи із розрахунків можна зробити висновок про доцільність науково-технічної роботи у рамках цієї теми. Успішна комерціалізація призведе до значних успіхів на ринку та забезпечення інвестора високим рівнем прибутків у досить короткі терміни.

ВИСНОВКИ

Магістерська кваліфікаційна робота присвячена аналізу та передбаченню добового генерування електроенергії фотоелектростанцією з урахуванням впливу сонячної активності.

В роботі здійснено аналіз предметної області, розглянуто наявність існуючих проблем, їх актуальність та шляхи вирішення. Досліджено основні методи, що використовуються в машинному навчанні для побудови моделей прогнозування та передбачення. Розглянуто аналоги, деякі з яких вже мали практичне застосування. Обґрунтована необхідність нового дослідження.

Здійснено аналіз вхідних даних та їх попередня обробка. Побудовані та проаналізовані добові графіки генерації електроенергії та сонячної радіації, визначена наявність кореляції між цими показниками. Проведений розвідувальний аналіз даних, в результаті якого побудовано статистичний звіт щодо показників датасету з візуалізацією кореляційних матриць Пірсона, Спірмена, Кендалла і Phik. За допомогою матриць визначена 93% кореляція між цільовою змінною та фактором впливу.

Побудовано декілька моделей передбачення на основі методів: `DecisionTreeRegressor`, `RandomForestRegressor`, `LightGBMRegressor`. Проведено їх навчання та перевірка точності їх передбачення, що супроводжувалося відповідними візуальними діаграмами.

По результатах дослідження обрано найкращу модель, якою виявилась модель, що побудована на основі метода `RandomForestRegressor` з точністю передбачення 0.99277. Така модель здатна з великою точністю передбачати добову генерацію електроенергії, що буде мати вагомий економічний ефект для фотоелектричної станції.

В розділі присвяченому економічній частині, проведені економічні розрахунки, що підтверджують великий комерційний потенціал розробки. Оціночний показник комерційного потенціалу становить 38 балів, що свідчить про значні можливості для успішної комерціалізації науково-технічного

рішення. Витрати на розробку покриваються за 1,34 року, що в свою чергу свідчить про високу комерційну привабливість розробки та швидке повернення грошей інвестора з подальшим поступовим збільшенням прибутку.

За домовленістю з адміністрацією однієї з фотоелектричних станцій у Вінницькій області, результат магістерської кваліфікаційної роботи буде випробуваний в реальних умовах з подальшою оцінкою його ефективності.

Результати роботи доповідались на ЛІІ Науково-технічній конференції факультету інтелектуальних інформаційних технологій та автоматизації ВНТУ (м. Вінниця, 2023 р.).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Форкун Я. Б., Шкурпела О.О. Сонячна теплоенергетика, конспект лекцій. Харків: ХНУМГ ім.О.М.Бекетова, 2020. 63 с.
2. Кудря С.О., Нетрадиційні та відновлювані джерела енергії. Київ: Національний технічний університет України («КПІ»), 2012. 240 с.
3. Мхитарян Н.М., Енергетика нетрадиційних та відновлюваних джерел. Київ: Наукова думка, 1999. 158 с.
4. Бартецький А.А., Гнатюк М.Ю., Кручай І.С. Передбачення погодинної генерації електроенергії ФЕС з врахуванням прогнозу хмарності // Матеріали ЛІІ Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації ВНТУ, Вінниця, 2023. URL: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2023/paper/view/18504/15353>. 11.2023
5. Solar Power Plant. URL: <https://www.repsol.com/en/energy-and-the-future/future-of-the-world/solar-power-plant/index.cshtml>. 11.2023
6. Титко Р., Відновлювальні Джерела Енергії (досвід Польщі для України): навч. посібник / за ред. В. М. Калініченко. Варшава: OWG, 2010. 294 с.
7. Шкляр В.И., Дубровская В.В., Карпенко Д.С. Використання енергії сонця фотоелектричними системами. Київ: Національний технічний університет України («КПІ»), 2009. 218 с.
8. Закон України про "Про ринок електричної енергії" від 13.04.2017 № 2019-VIII. URL: <http://zakon5.rada.gov.ua/laws/show/2019-19>. 11.2023
9. Forecast vs. Prediction in HR: What's the Difference? URL: <https://www.visier.com/blog/forecast-vs-prediction-in-hr>. 11.2023
10. Аналіз метеопараметрів для погодинного прогнозування виробітку електроенергії фотовольтаїчними електростанціями на добу вперед / Лежнюк П. Д., Комар В. О., Кравчук С. В., Дідіченко Є. С. Вінниця: ВНТУ, 2017. 46 с.
11. M. Bouzerdoum, A. Mellit, A. Massi Pavan. A hybrid model for short-term power forecasting of a small-scale grid-connected photovoltaic plant. URL:

<https://www.sciencedirect.com/science/article/abs/pii/S0038092X14001522>.

11.2023

12. M.Q. Raza, M. Nadarajah, C. Ekanayake. Review On recent advances in PV output power forecast. URL:

<https://www.sciencedirect.com/science/article/abs/pii/S0038092X16302547>.

11.2023

13. M. Abuella, B. Chowdhury. Solar power forecasting using artificial neural networks. NAPS, 2015. URL: https://www.researchgate.net/profile/Mohamed-Abuella/publication/282650839_Solar_Power_Forecasting_Using_Artificial_Neural_Networks/links/5616486208ae0f2140069fb0/Solar-Power-Forecasting-Using-Artificial-Neural-Networks.pdf. 11.2023

14. Data Science. URL: <https://coursehunter.net/course/data-science-uroven1-instrumenty-i-tehnologii>. 11.2023

15. Мокін В. Б. Heart Disease - Automatic AdvEDA & FE & 20 models. URL: <https://www.kaggle.com/vbmokin/heart-disease-comparison-of-20-models>.

11.2023

16. Наївний баєсів класифікатор URL: https://uk.wikipedia.org/wiki/Наївний_баєсів_класифікатор. 11.2023

17. Decision Tree URL: <https://www.geeksforgeeks.org/decision-tree>. 11.2023

18. Rafał Weron. Modeling and Forecasting Electricity Loads and Prices. Wrocław University of Technology, 2006. URL: <http://prac.im.pwr.wroc.pl/~hugo/RePEc/wuu/hsbook/HSBook0601.pdf>. 11.2023

19. Online 24-h solar power forecasting based on weather type classification using artificial neural network / C. Chen., S. Duan., T. Cai., B. Liu. URL: <https://www.scienceopen.com/document?vid=c1cc79d4-88f2-4935-9d7d-b6eb736ed057>. 11.2023

20. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator / F. Almonacid., P. J. Perez-Higueras., E. F. Fernandez., L. Hontoria. URL:

https://www.researchgate.net/publication/277650408_A_methodology_based_on_dynamic_artificial_neural_network_for_short-term_forecasting_of_the_power_output_of_a_PV_generator. 11.2023

21. A. Vaz, B. Elsinga, M. van Sark. An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht. The Netherlands, 2014. URL: <https://ideas.repec.org/a/eee/renene/v85y2016icp631-641.html>. 11.2023

22. Pattern sequence neural network for solar power forecasting / Y. Lin, I. Koprinska, M. Rana, A. Troncoso. URL: https://www.researchgate.net/publication/337779342_Pattern_Sequence_Neural_Network_for_Solar_Power_Forecasting. 11.2023

23. K. Wang., X. Qi., H. Liu. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. Appl. Energy, 2019. URL: https://econpapers.repec.org/article/eeeappene/v_3a251_3ay_3a2019_3ai_3ac_3a30.htm. 11.2023

24. Zhang, J., Verschae, R., Nobuhara J.F. Deep photovoltaic nowcasting. Sol. Energy, 2018. URL: <https://www.scinapse.io/papers/2963986342>

25. Load forecasting via deep neural networks. Procedia Comput. Sci., 2017. URL: <https://isiarticles.com/bundles/Article/pre/pdf/155400.pdf>

26. Zhengwei Huang, Jin Huang, Jintao Min. Short-Term Photovoltaic Power Prediction Based on Feature Matching. URL: <https://www.mdpi.com/1996-1073/15/20/7806>. 11.2023

27. Kaggle: All you need to know about this platform. URL: <https://datascientest.com/en/kaggle-all-about-this-platform>. 11.2023

28. What Is Kaggle? How to Compete in Kaggle Competitions. URL: <https://builtin.com/data-science/what-is-kaggle>. 11.2023

29. Python. URL: <https://aws.amazon.com/ru/what-is/python/>. 11.2023

30. A. Bandyopadhyay. Hands-On GPU Computing with Python. Packt

Publishing, 2019. URL: <https://www.packtpub.com/product/hands-on-gpu-computing-with-python/9781789341072>. 11.2023

31. Ває І., J. Kim. Reliability Evaluation of Distributed Generation. URL: <https://ietresearch.onlinelibrary.wiley.com/doi/pdfdirect/10.1049/iet-gtd.2015.0669>

32. Регресійний аналіз. URL: https://uk.wikipedia.org/wiki/Регресійний_аналіз. 11.2023

33. Bishop C.M. Pattern Recognition and Machine Learning. NY: Springer, 2006. URL: <http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf>. 11.2023

34. Бібліотека Pandas. URL: <https://uk.wikipedia.org/wiki/Pandas> . 11.2023

35. Chowdhury A. A. Reliability Modeling of Distributed Generation in Conventional Distribution Systems Planning and Analysis. URL: https://www.researchgate.net/publication/3171659_Reliability_Modeling_of_Distributed_Generation_in_Conventional_Distribution_Systems_Planning_and_Analysis. 11.2023

36. Sutskever I. Sequence to sequence learning with neural networks. 2014. URL: <https://theses.hal.science/tel-03884481v1/document>. 11.2023

37. Bruce E. Pease. Pease Leading Intelligence Analysis. CQ Press. URL: https://books.google.com.ua/books/about/Leading_Intelligence_Analysis.html?id=gqhkDwAAQBAJ&redir_esc=y. 11.2023

38. Pearson correlation coefficient: Definition, formula & calculation. URL: <https://www.questionpro.com/blog/pearson-correlation-coefficient/>. 11.2023

39. Spearman's Rank Correlation: The Definitive Guide To Understand. URL: <https://www.simplilearn.com/tutorials/statistics-tutorial/spearman's-rank-correlation>. 11.2023

40. Kendall Rank Correlation Explained. URL: <https://towardsdatascience.com/kendall-rank-correlation-explained-dee01d99c535>. 11.2023

41. Phik (ϕ_k) — get familiar with the latest correlation coefficient. URL: <https://towardsdatascience.com/phik-k-get-familiar-with-the-latest-correlation-coefficient-9ba0032b37e7>. 11.2023

42. Козловський В. О., Лесько О. Й., Кавецький В. В. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт: уклад. Вінниця: ВНТУ, 2021. 42 с.

Додаток А
(обов'язковий)

Міністерство освіти і науки України
Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

_____ д.т.н., проф. Віталій МОКІН

«___» _____ 2023 р.

ТЕХНІЧНЕ ЗАВДАННЯ

на комплексну магістерську кваліфікаційну роботу
АНАЛІЗ ТА ПЕРЕДБАЧЕННЯ ДОБОВОГО ГЕНЕРУВАННЯ
ЕЛЕКТРОЕНЕРГІЇ ФОТОЕЛЕКТРОСТАНЦІЄЮ. ЧАСТИНА 3. АНАЛІЗ
ВПЛИВУ СОНЯЧНОЇ АКТИВНОСТІ
08-34.МКР.002.02.000 ТЗ

Керівник: д.т.н., проф. каф. САІТ

_____ Олександр МОКІН

«___» _____ 2023 р.

Розробив: студент гр. 2ІСТ-22м

_____ Ігор КРУЧАЙ

«___» _____ 2023 р.

1. Підстава для проведення робіт.

Підставою для виконання роботи є наказ № __ по ВНТУ від «__» _____ 2023 р., та індивідуальне завдання на МКР, затверджене протоколом № __ засідання кафедри САІТ від «__» _____ 2023 р.

2. Джерела розробки:

1) Kaggle Dataset „EnergyAndSolarRadiation” <https://www.kaggle.com/datasets/igorkruchay/energyandsolarradiation>, наданий з дозволу адміністрації фотоелектричної станції, розташованої у Вінницькій області.

3. Мета і призначення роботи.

Підвищення точності передбачення добового генерування електроенергії фотоелектростанцією за рахунок використання методів машинного навчання з урахуванням впливу сонячної активності.

4. Вихідні дані для проведення робіт:

Набір даних моніторингу генерації електроенергії фотоелектричної станції, розташованої у Вінницькій області.

5. Методи дослідження.

1) Розвідувальний аналіз.

2) Передбачення даних.

6. Етапи роботи і терміни їх виконання:

а) Аналіз предметної області	21.09 – 30.09
б) Вибір оптимальних інформаційних технологій	01.10 – 09.10
в) Розроблення інформаційної технології	10.10 – 24.10
г) Тренування та тестування моделей машинного навчання	25.10 – 28.10
д) Економічна частина	29.10 – 03.11
е) Оформлення пояснювальної записки	04.11 – 25.11

7. Очікувані результати та порядок реалізації.

Передбачити генерацію електроенергії фотоелектричною станцією за вказаними даними: числом та місяцем року.

8. Вимоги до розробленої документації.

Текстова та ілюстративна частини роботи оформлені у відповідності до вимог «Методичних вказівок до виконання магістерських кваліфікацій-них робіт для студентів спеціальності 126 «Інформаційні системи та технології» (освітня програма «Інформаційні технології аналізу даних та зображень»).

9. Порядок приймання роботи

Публічний захист.....«19» _____ грудня _____ 2023 р.

Початок розробки «21» _____ вересня _____ 2023 р.

Граничні терміни виконання МКР «30» _____ листопада _____ 2023 р.

Розробив студент групи 2ІСТ-22м _____ Ігор КРУЧАЙ

Додаток Б
(обов'язковий)

Протокол перевірки магістерської кваліфікаційної роботи на наявність
текстових запозичень

Назва роботи: «Аналіз та передбачення добового генерування електроенергії фотоелектростанцією. Частина 3. Аналіз впливу сонячної активності»

Тип роботи: магістерська кваліфікаційна робота

Підрозділ: кафедра САІТ

Показники звіту подібності Unicheck

Оригінальність 92,11 %

Схожість 7,89 %

Аналіз звіту подібності (відмітити потрібне)

- Запозичення, виявлені у роботі, оформлені коректно і не містять ознак плагіату.
- Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і самостійності її автора. Роботу направити на розгляд експертної комісії кафедри.
- Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

Особа, відповідальна за перевірку


(підпис)

Сергій ЖУКОВ

Ознайомлені з повним звітом подібності, який був згенерований системою Unicheck щодо роботи.

Автор роботи


(підпис)

Ігор КРУЧАЙ

Керівник роботи


(підпис)

Олександр МОКІН

Додаток В
(ДОВІДКОВИЙ)

Лістинг програми

```

df_solar = pd.read_csv(
'/kaggle/input/energyandsolarradiation/EnergyAndSolarRadiation.csv', sep=';',
header=0)

# splitting base dataset into two: "training" 60% and "test" 40%
df = df_solar.copy()
X = df[['TOTAL_DAYS','TOTAL_TIME','SOLAR_RADIATION']]
y = df[['ENERGY']]
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=.4,random_state=0)

print(X_train.shape, X_test.shape)

# splitting "train" dataset into two: "validation" 20% and "train" 80%
X_train_all = X_train.copy()
y_train_all = y_train.copy()
X_train, X_valid, y_train, y_valid = train_test_split(X_train_all, y_train_all,
test_size=0.2, random_state=0)

# Creation the dataframe with the resulting score of all models
result = pd.DataFrame({'model': [
'DecisionTreeRegressor',
'RandomForestRegressor',
'LightGBMRegressor'
],
'train_score': 0,
'valid_score': 0})

##### Decision Tree Regressor #####
from sklearn.tree import DecisionTreeRegressor
dtr = DecisionTreeRegressor()
dtr.fit(X_train,y_train['ENERGY'])

# Prediction for training dataset
y_pred_dtr = dtr.predict(X_train)

# Accuracy of model
r2_score_dtr = round(r2_score(y_train['ENERGY'], y_pred_dtr), 5)

```

```

# Save to result dataframe
result.loc[result['model'] == 'DecisionTreeRegressor', 'train_score'] = r2_score_dtr
print(f'DecisionTreeRegressor model, accuracy prediction for training dataset is:
{r2_score_dtr}')

# Prediction for validation dataset
y_val_dtr = dtr.predict(X_valid)

# Accuracy of model
r2_score_dtr_valid = round(r2_score(y_valid['ENERGY'], y_val_dtr),5)

# Save to result dataframe
result.loc[result['model'] == 'DecisionTreeRegressor', 'valid_score'] =
r2_score_dtr_valid

print(f'DecisionTreeRegressor model, accuracy prediction for validation dataset is:
{r2_score_dtr_valid}')

##### Random Forest Regressor #####
from sklearn.ensemble import RandomForestRegressor
rfr = RandomForestRegressor()

# Training model
rfr.fit(X_train,y_train['ENERGY'])

# Prediction for training dataset
y_pred_rfr = rfr.predict(X_train)

# Accuracy of model
r2_Score_rfr = round(r2_score(y_train['ENERGY'], y_pred_rfr), 5)

# Save to result dataframe
result.loc[result['model'] == 'RandomForestRegressor', 'train_score'] =
r2_Score_rfr

print(f'RandomForestRegressor model, accuracy prediction for training dataset is:
{r2_Score_rfr}')

# Prediction for validation dataset
y_val_rfr = rfr.predict(X_valid)

# Accuracy of model
r2_score_rfr_valid = round(r2_score(y_valid['ENERGY'], y_val_rfr),5)

```

```

# Save to result dataframe
result.loc[result['model'] == 'RandomForestRegressor', 'valid_score'] =
r2_score_rfr_valid

print(f'RandomForestRegressor model, accuracy prediction for validation dataset
is: {r2_score_rfr_valid}')

##### LGBMClassifier #####
import lightgbm as lgb
from lightgbm import LGBMClassifier
train_data = lgb.Dataset(X_train, label=y_train['ENERGY'])
params = {
    'num_leaves': 50,
    'learning_rate': 0.05,
    'metric': 'mae',
}

# Training model
model_lgb = lgb.train(params, train_data, num_boost_round=1000)

# Prediction for training dataset
y_pred_lgb = model_lgb.predict(X_train)

# Accuracy of model
r2_Score_lgb = round(r2_score(y_train['ENERGY'], y_pred_lgb), 5)

# Save to result dataframe
result.loc[result['model'] == 'LightGBMRegressor', 'train_score'] = r2_Score_lgb

print(f'LightGBMRegressor model, accuracy prediction for training dataset is:
{r2_Score_lgb}')

# Prediction for validation dataset
y_val_lgb = model_lgb.predict(X_valid)

# Accuracy of model
r2_score_lgb_valid = round(r2_score(y_valid['ENERGY'], y_val_lgb), 5)

# Save to result dataframe
result.loc[result['model'] == 'LightGBMRegressor', 'valid_score'] =
r2_score_lgb_valid

```

```
print(f'LightGBMRegressor model, accuracy prediction for validation dataset is:  
{r2_score_lgb_valid}')
```

```
# Filtering models with big difference between training and validation datasets (dif  
> 10%)
```

```
result_best = result[(result['train_score'] - result['valid_score']).abs() < 10]  
result_best.sort_values(by=['valid_score', 'train_score'], ascending=False)
```

```
# Choosing the best model
```

```
result_best = result.nlargest(1, 'valid_score')  
result_best
```

Додаток Г
(обов'язковий)

ІЛЮСТРАТИВНА ЧАСТИНА

АНАЛІЗ ТА ПЕРЕДБАЧЕННЯ ДОБОВОГО ГЕНЕРУВАННЯ
ЕЛЕКТРОЕНЕРГІЇ ФОТОЕЛЕКТРОСТАНЦІЄЮ. ЧАСТИНА 3. АНАЛІЗ
ВПЛИВУ СОНЯЧНОЇ АКТИВНОСТІ

Нормоконтроль: к.т.н., доцент

_____ Сергій ЖУКОВ

«___» _____ 2023 р.

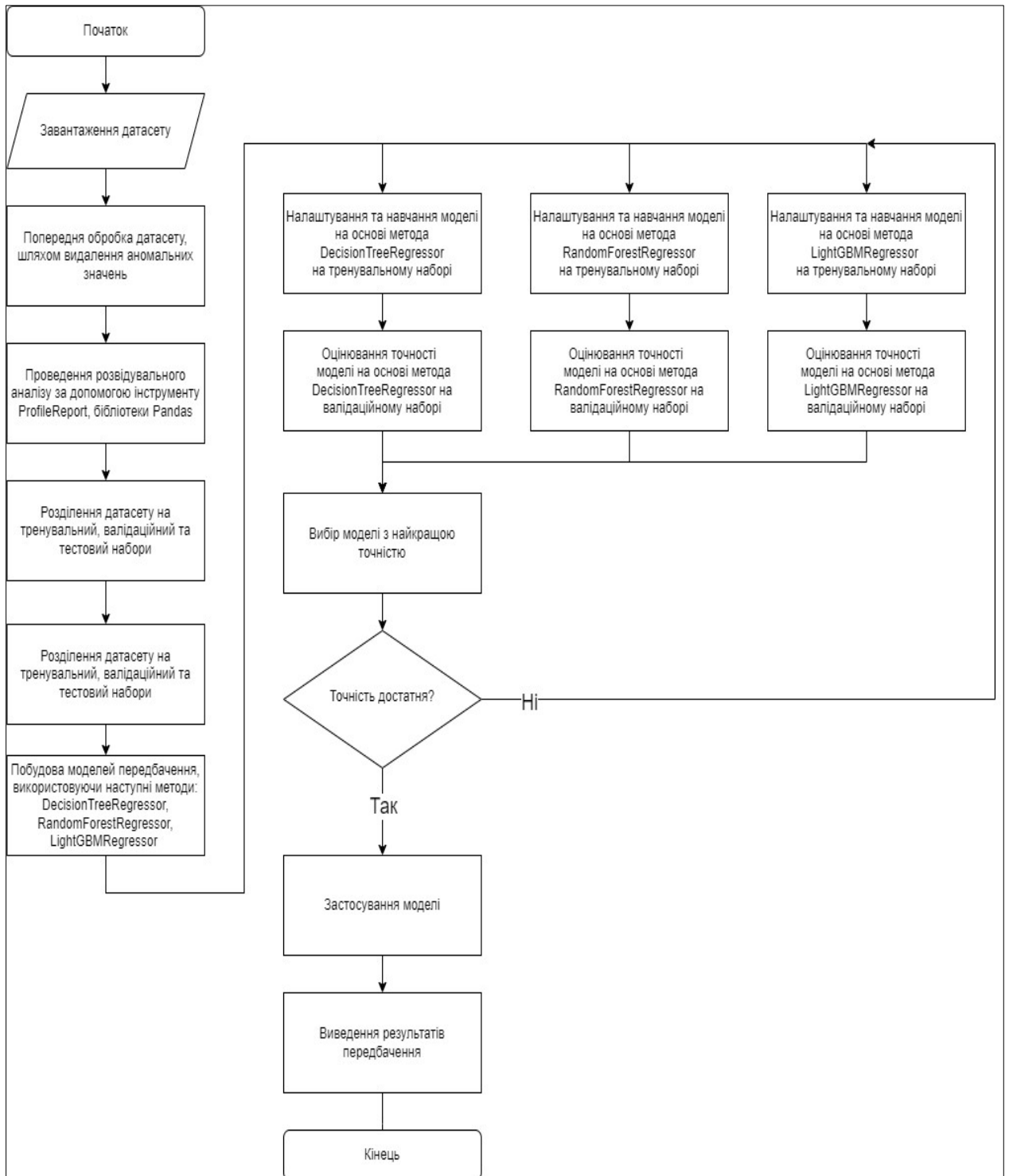


Рисунок Г.1 – Блок схема алгоритму інформаційної технології передбачення добового генерування електроенергії фотоелектростанцією з урахуванням сонячної активності

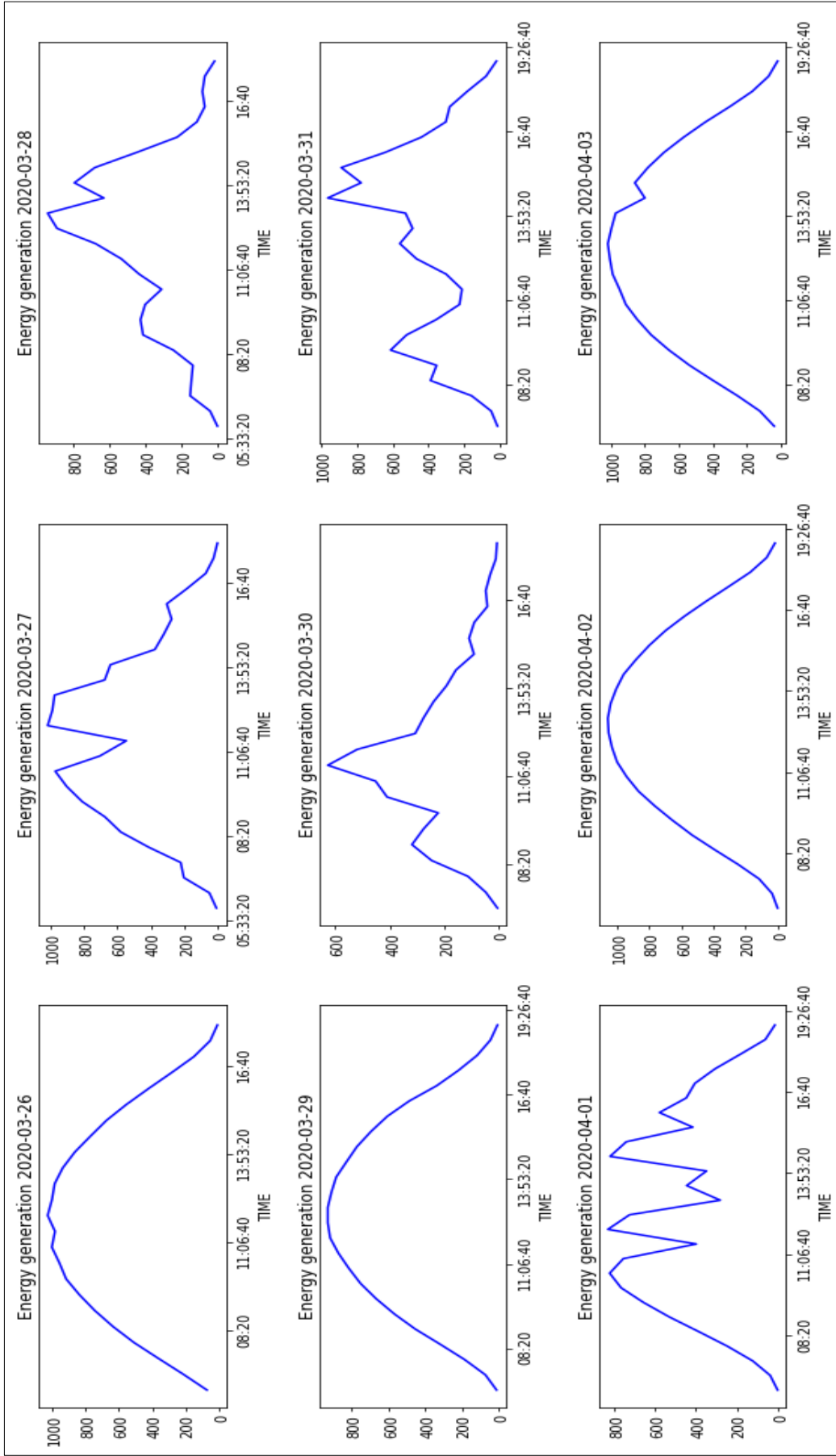


Рисунок Г.2 – Добові графіки генерації електроенергії

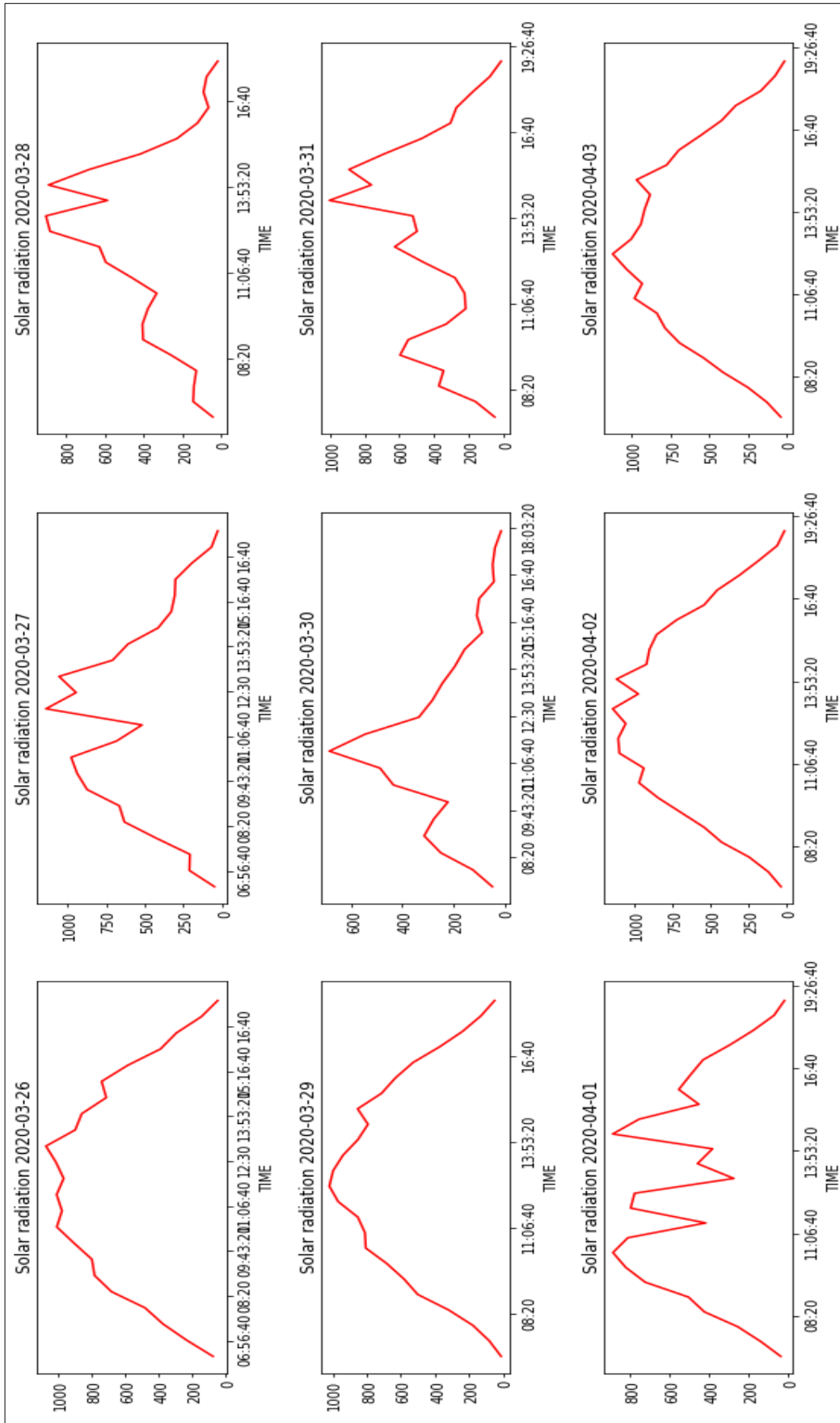


Рисунок Г.3 – Добові графіки сонячної активності

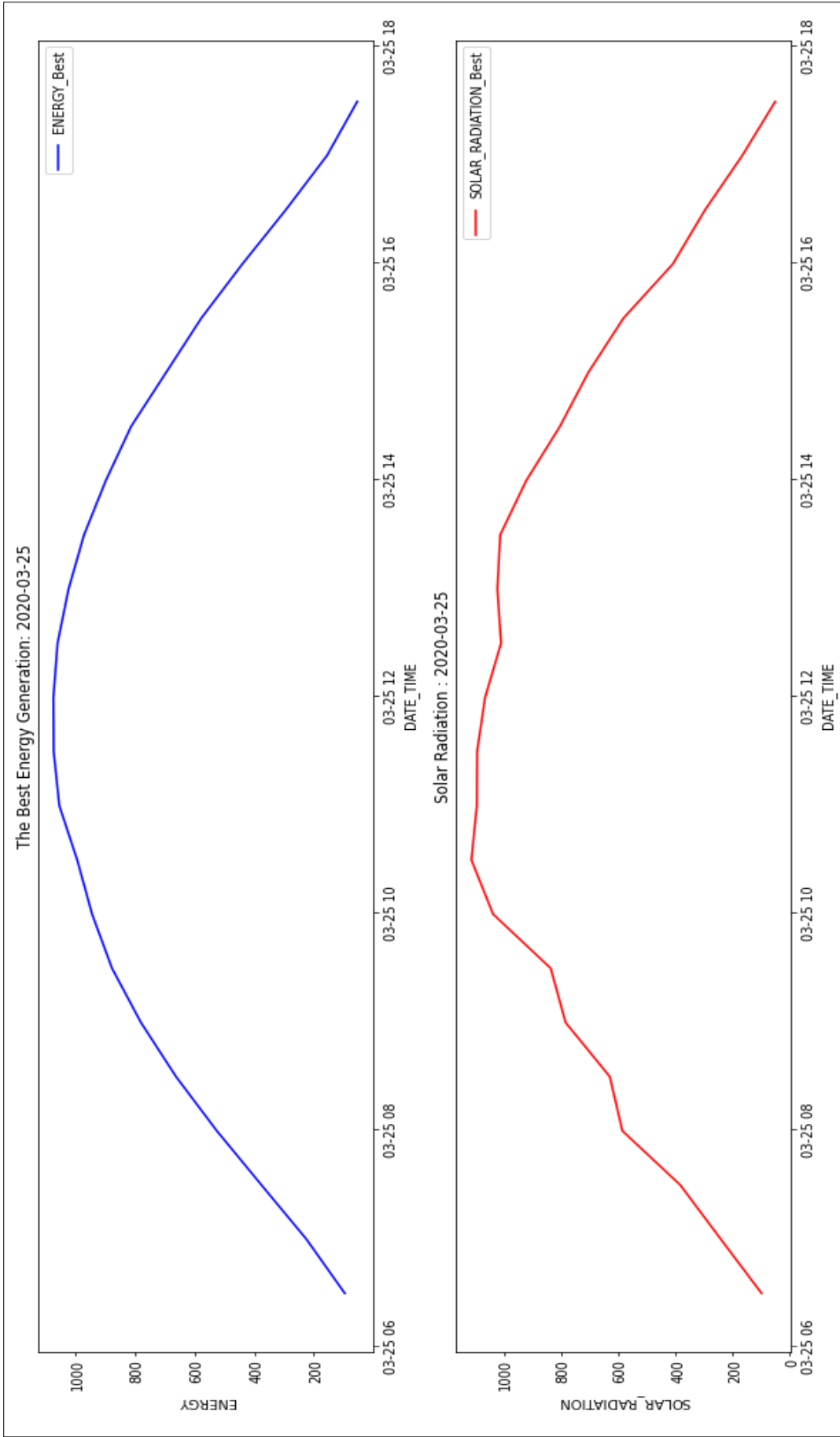


Рисунок Г.4 – Графіки генерації електроенергії та фіксації сонячної активності 25 березня 2020 року

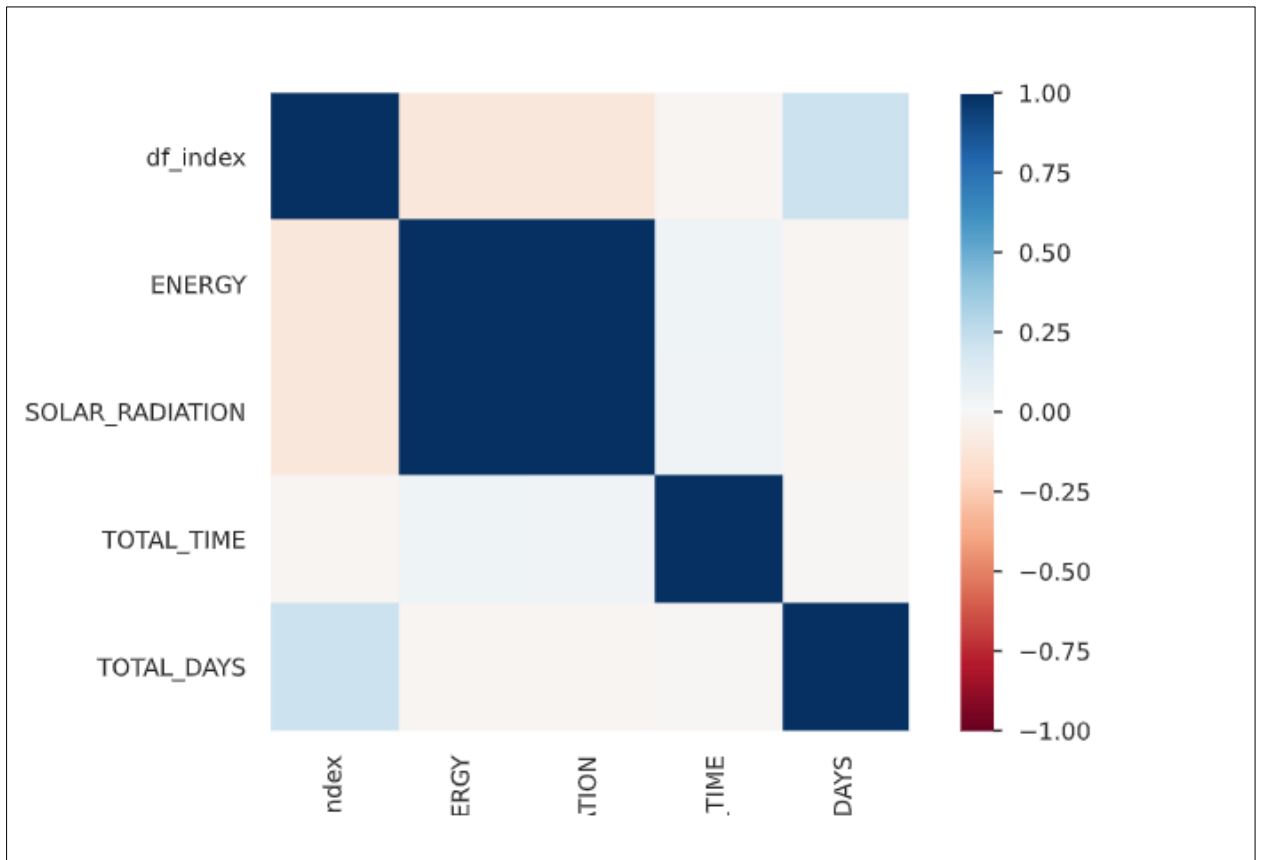


Рисунок Г.5 – Матриця кореляції Спірмена

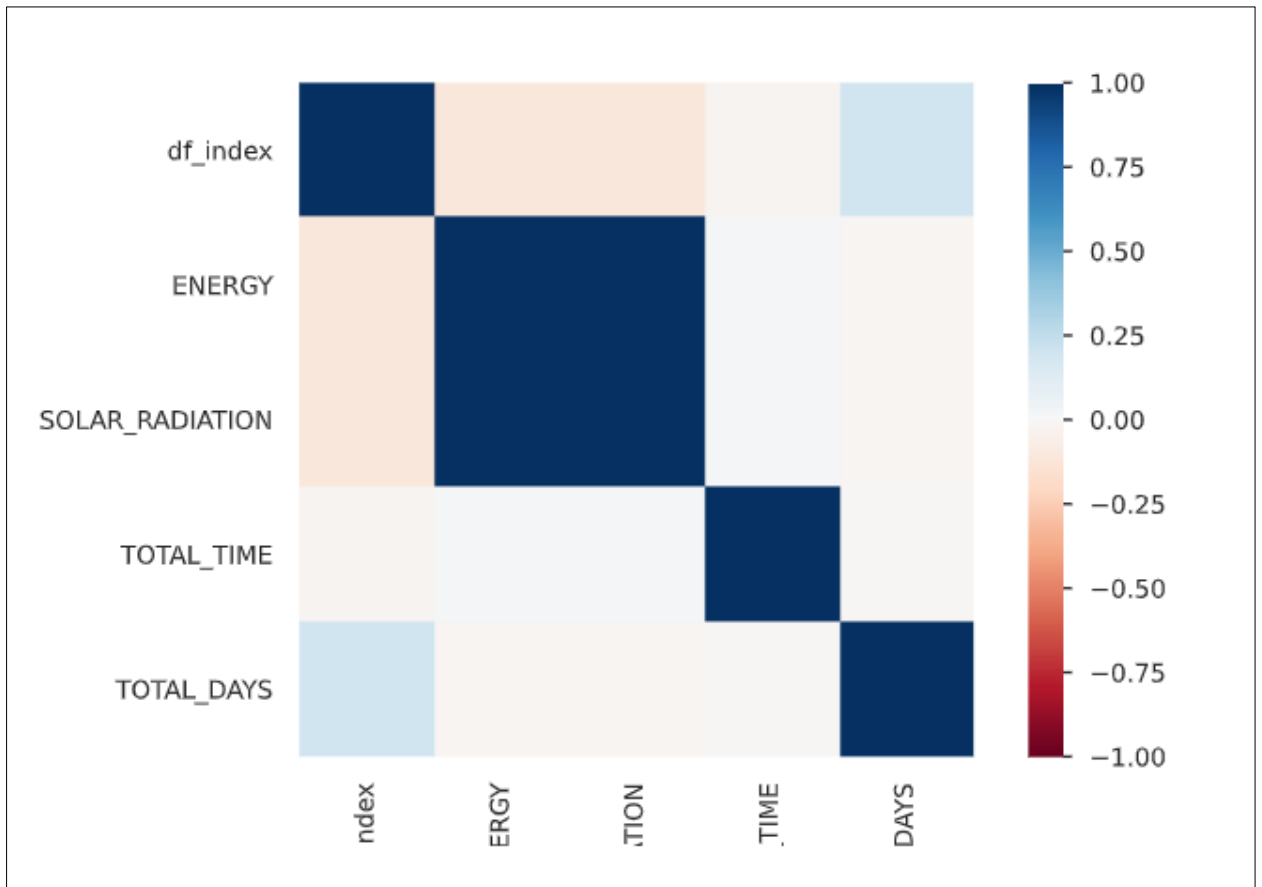


Рисунок Г.6 – Матриця кореляції Пірсона

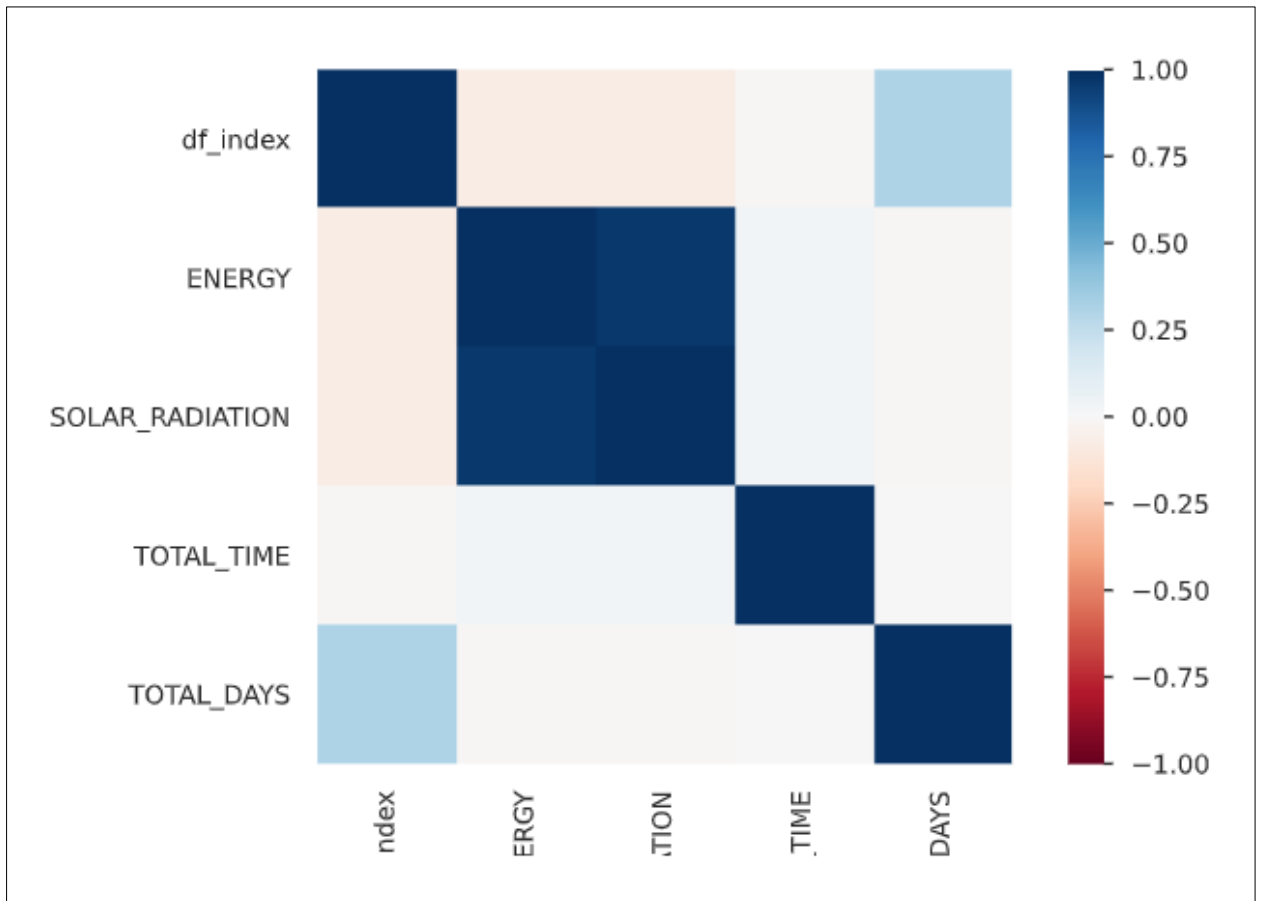


Рисунок Г.7 – Матриця кореляції Кендала

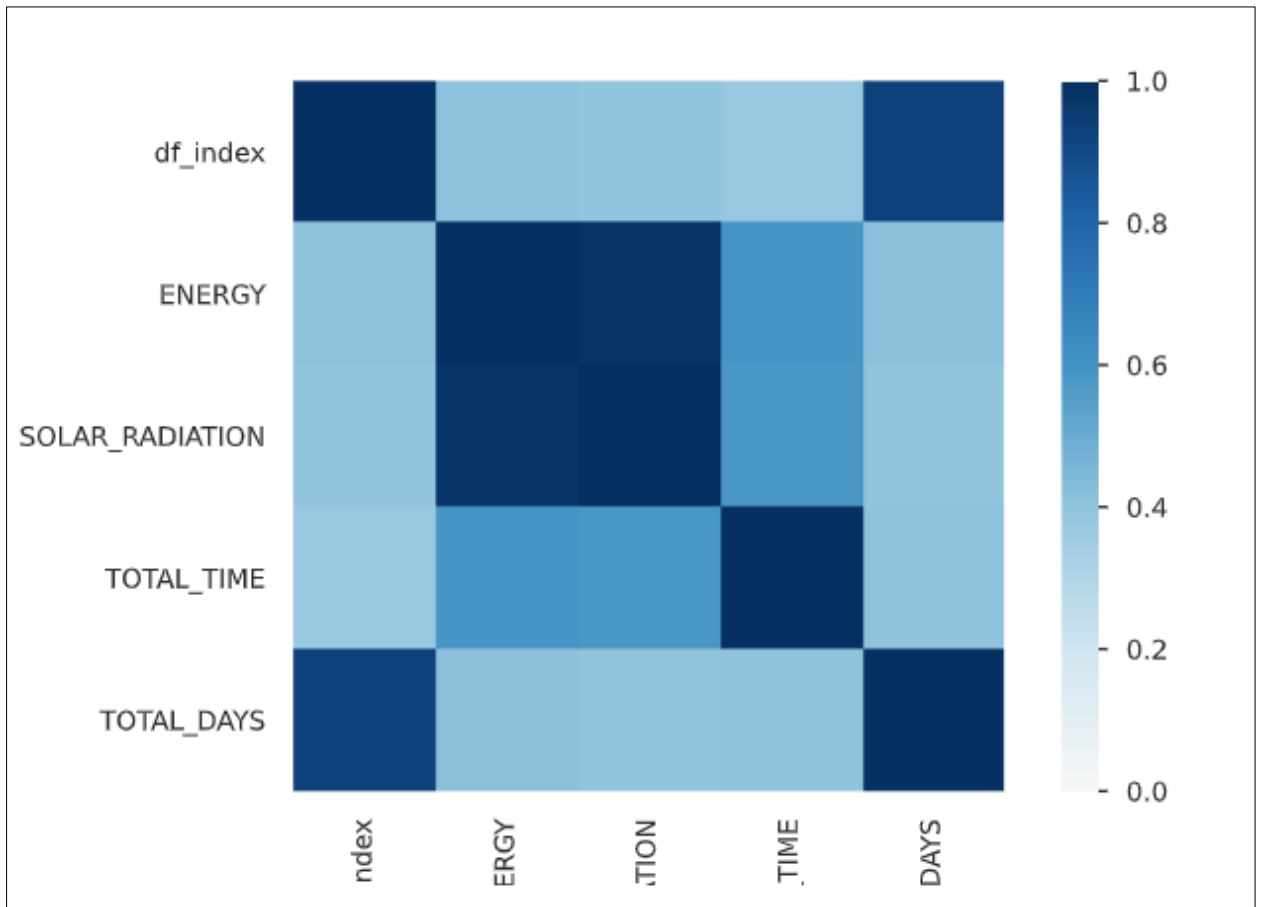


Рисунок Г.8 – Матриця кореляції Phik

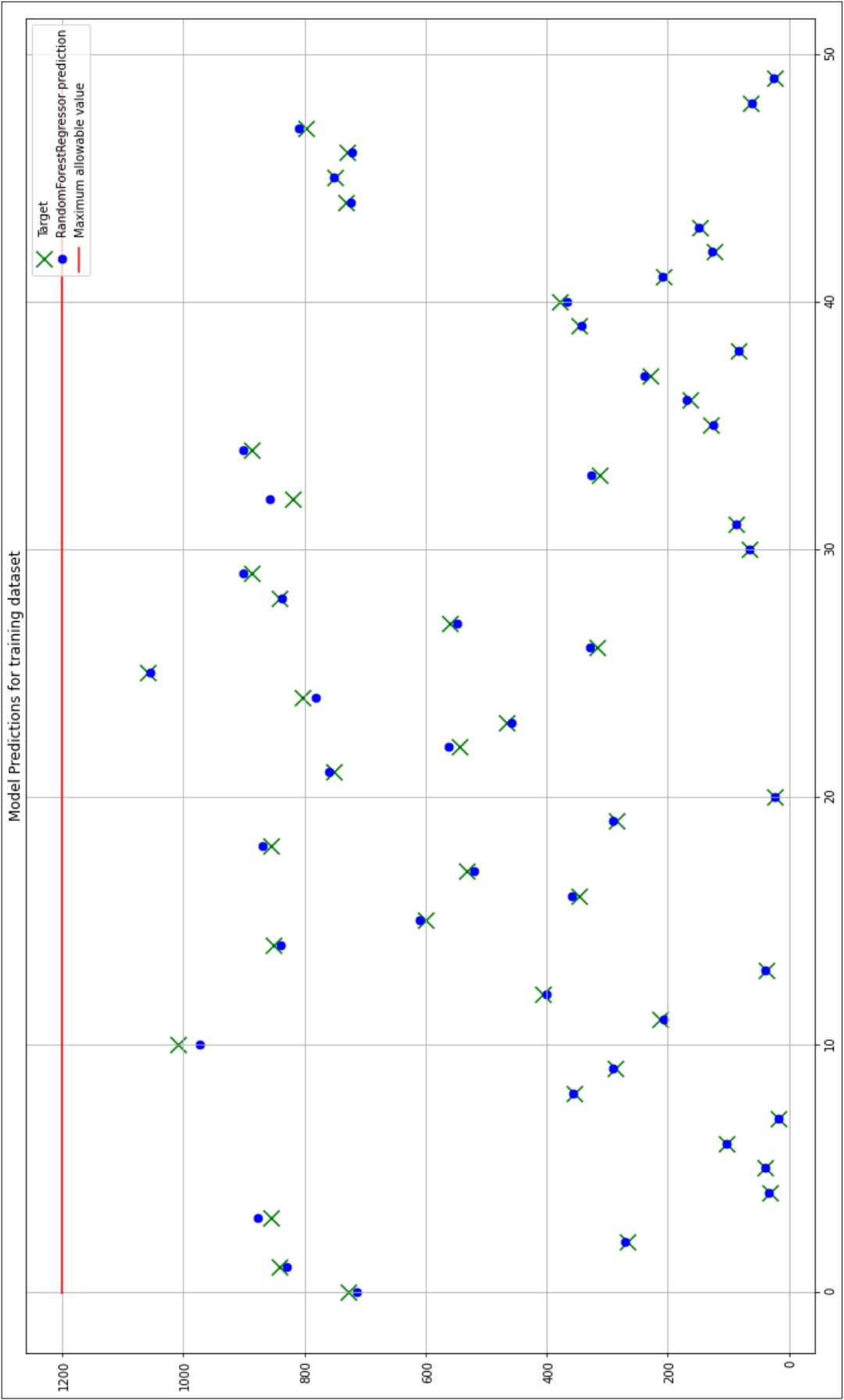


Рисунок Г.9 – Візуалізація передбачення моделі на основі метода RandomForrestRegressor для тренувального набору даних

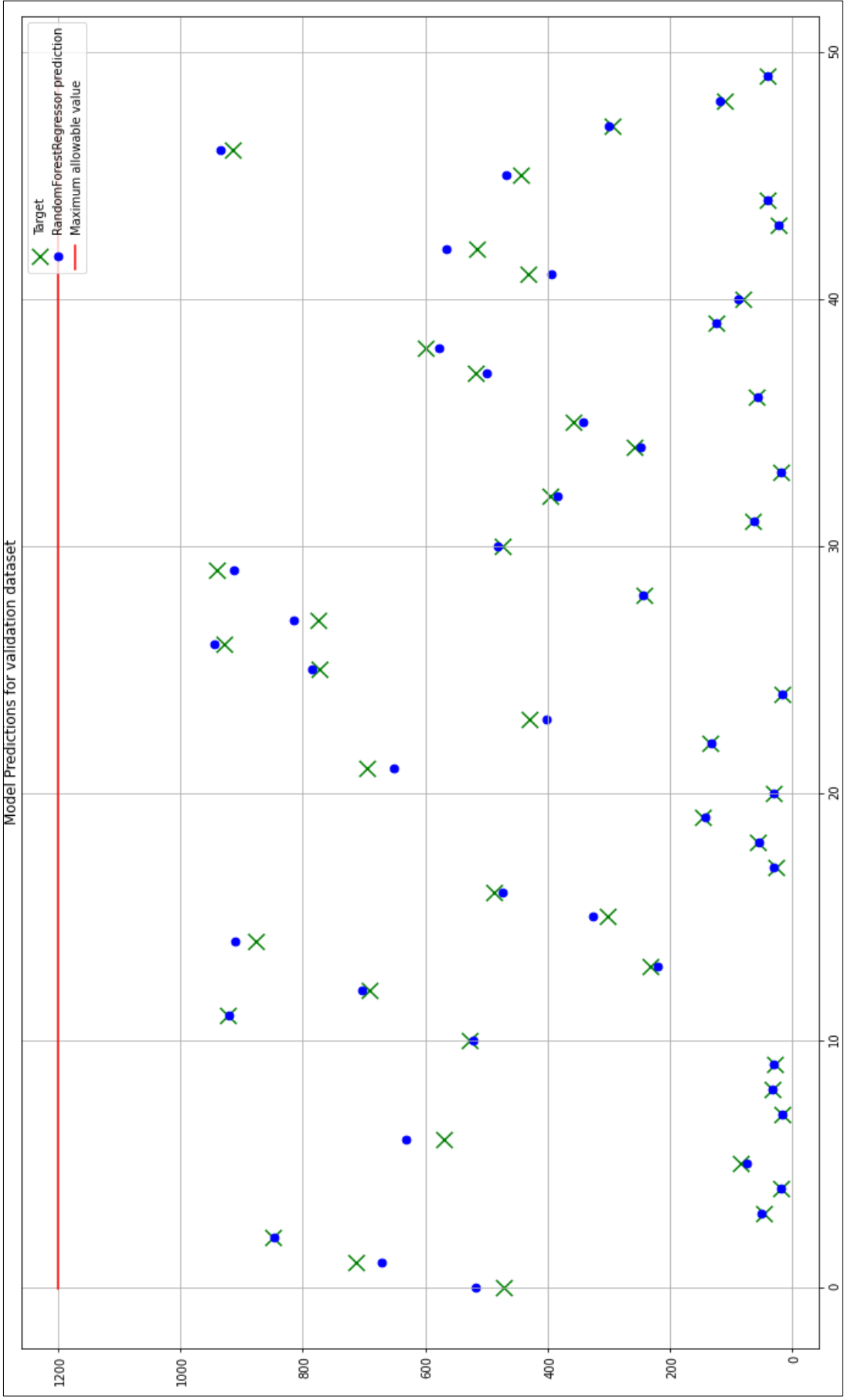


Рисунок Г.10 – Візуалізація передбачення моделі на основі метода RandomForestRegressor для валідаційного набору даних

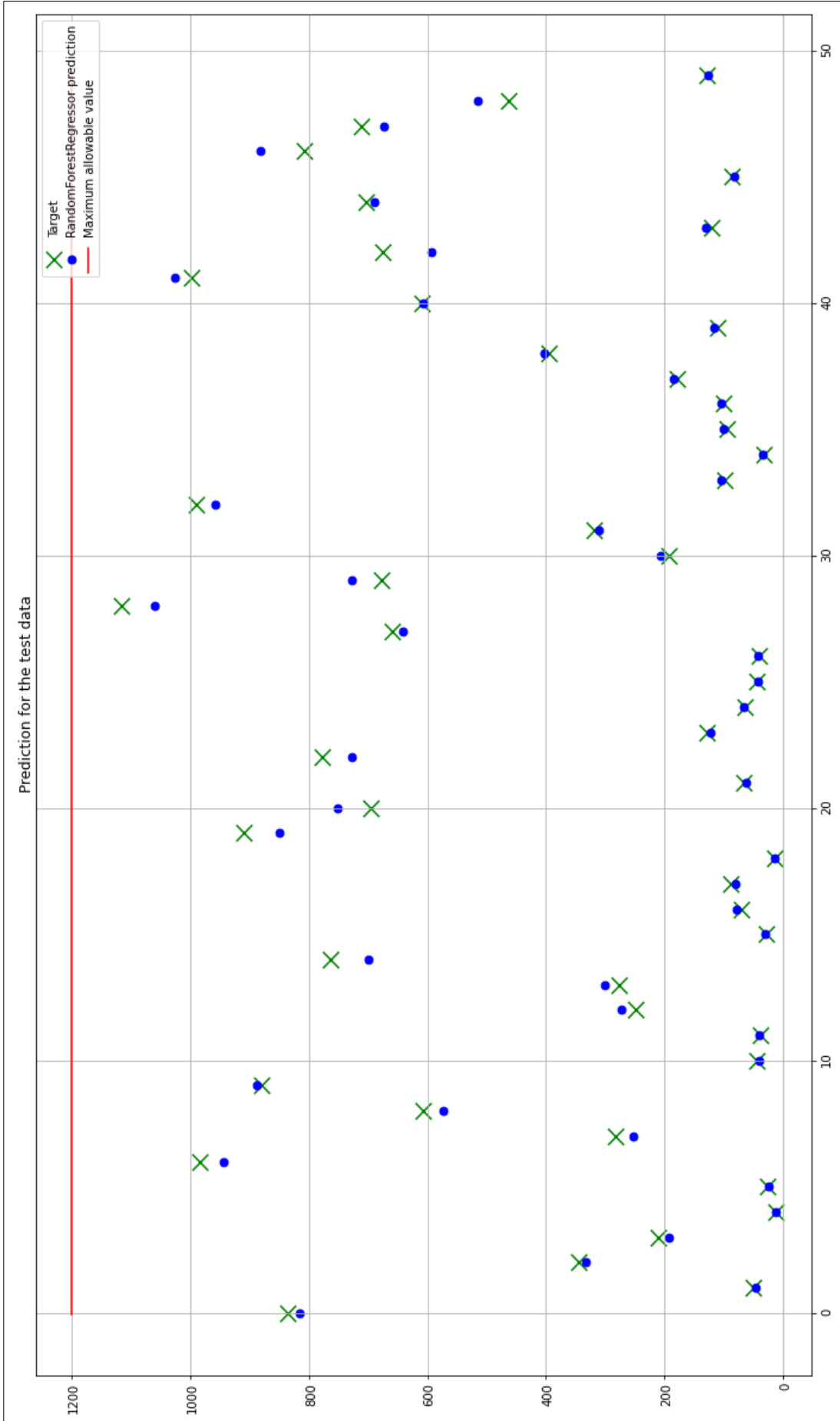


Рисунок Г.11 – Візуалізація передбачення моделі на основі метода RandomForestRegressor для тестового набору даних

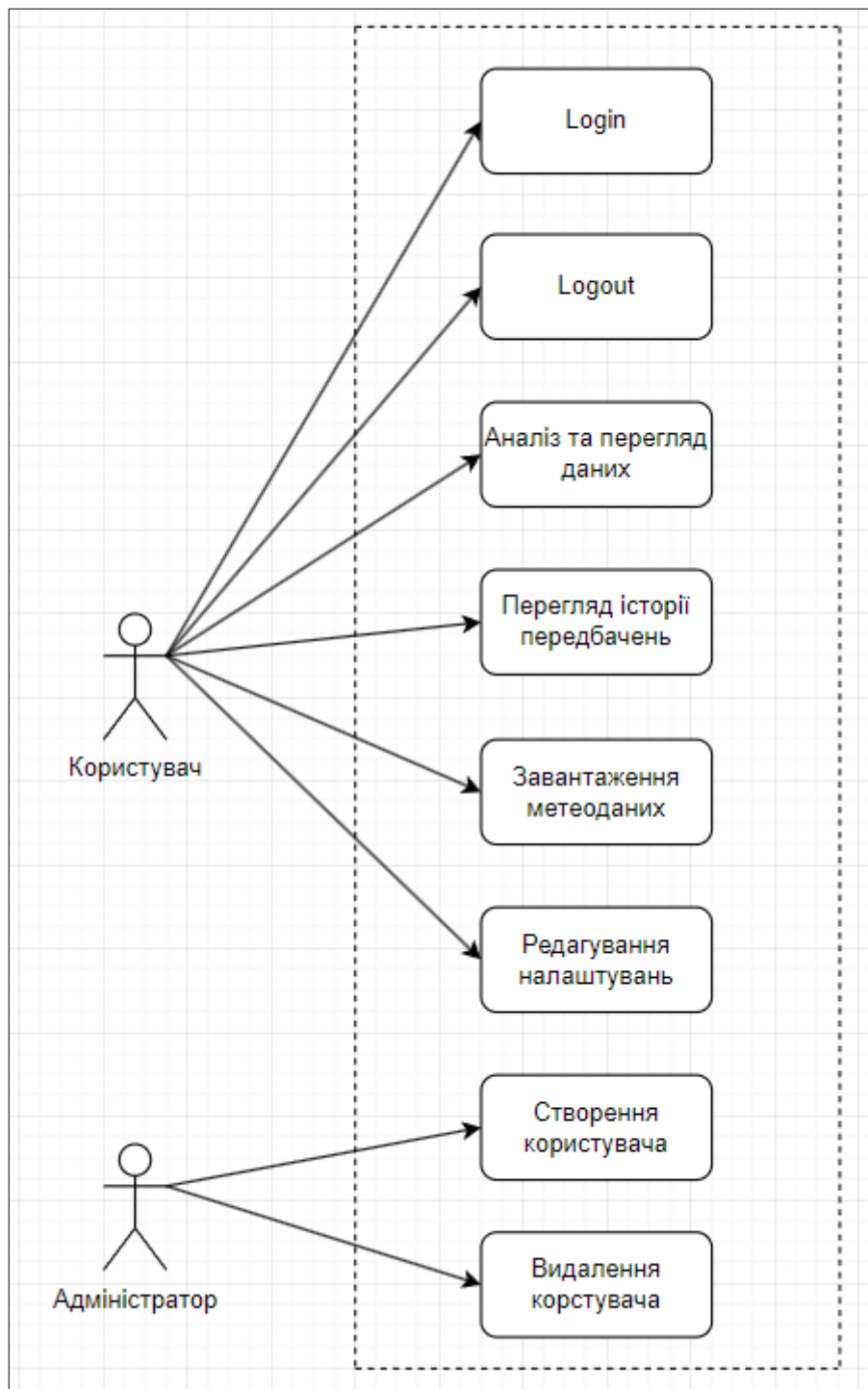


Рисунок Г.12 – Use-case діаграма використання інформаційної технології