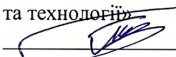


Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій

Комплексна магістерська кваліфікаційна робота на тему:

**АНАЛІЗ ТА ПЕРЕДБАЧЕННЯ ДОБОВОГО ГЕНЕРУВАННЯ ЕЛЕКТРОЕНЕРГІЇ
ФОТОЕЛЕКТРОСТАНЦІЄЮ. ЧАСТИНА 2. АНАЛІЗ ВПЛИВУ ХМАРНОСТІ**

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


 Михайло ГНАТЮК

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 – другий (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» 07 2023 року

**ЗАВДАННЯ
НА КОМПЛЕКСНУ МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ
СТУДЕНТУ**

Гнатюку Михайлу Юрійовичу

1. Тема роботи: «Аналіз та передбачення добового генерування електроенергії фотоелектростанцією. Частина 2. Аналіз впливу хмарності»
керівник роботи: Олександр МОКІН, д.т.н., проф. каф. САІТ
затверджені наказом ВНТУ від «18» 07 2023 року № 247
2. Термін подання студентом роботи 30.11.2023
3. Вихідні дані до роботи:
Kaggle Dataset „Energy” <https://www.kaggle.com/datasets/hnatyukmu/energy>,
наданий з дозволу адміністрації фотоелектростанції, розташованої на території Вінницької області.
4. Зміст текстової частини:
 - 1) Аналіз предметної області;
 - 2) Вибір оптимальних налаштувань для розв’язання поставленої задачі;
 - 3) Системний аналіз та передбачення добового генерування електроенергії фотоелектростанцією;
 - 4) Економічна частина.
5. Перелік ілюстративного матеріалу:
 - 1) Добові графіки генерації електроенергії;
 - 2) Добові графіки зміни хмарності;
 - 3) Графіки кореляційних матриць між ознаками;
 - 4) Блок-схема алгоритму інформаційної технології;
 - 5) Вибір оптимальної моделі передбачення;
 - 6) Графіки передбачень для різних наборів даних;
 - 7) Use-Case діаграма розробленої інформаційної технології

6. Консультанти розділів роботи

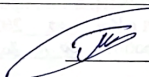
Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв
4	Наталія БУРЕННІКОВА, д.е.н., проф. каф. ЕПВМ	09/10 	09/11 

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	<i>Бен</i>
2	Вибір оптимальних інформаційних технологій	01.10.23	09.10.23	<i>Бен</i>
3	Розроблення інформаційної технології	10.10.23	24.10.23	<i>Бен</i>
4	Тренування та тестування моделей машинного навчання	25.10.23	28.10.23	<i>Бен</i>
5	Економічна частина	29.10.23	03.11.23	<i>Бен</i>
6	Оформлення матеріалів до захисту МКР	04.11.23	25.11.23	<i>Бен</i>

Студент



Михайло ГНАТЮК

Керівник роботи



Олександр МОКІН

АНОТАЦІЯ

УДК 004.08

Гнатюк М.Ю. Аналіз та передбачення добового генерування електроенергії фотоелектростанцією. Частина 2. Аналіз впливу хмарності. Комплексна магістерська кваліфікаційна робота зі спеціальності 126 – інформаційні системи та технології, освітньо-професійна програма – інформаційні технології аналізу даних та зображень. Вінниця: ВНТУ, 2023. 98 с.

На укр. мові. Бібліогр.: 37 назв; рис.: 69; табл.: 8.

В магістерській кваліфікаційній роботі розроблено процес аналізу та передбачення добового генерування електроенергії фотоелектростанцією з урахуванням хмарності. Використано відкриті дані з щодобовим узагальненням. Проаналізовано основні закономірності. Побудовано ряд моделей з використанням технологій машинного навчання та вибрано оптимальну серед них. Здійснено передбачення генерування електроенергії фотоелектростанцією та виконано аналіз, щодо ефективності роботи оптимальної моделі.

Ілюстративна частина складається з 11 плакатів.

У розділі економічної частини розглянуто питання про доцільність розробки та впровадження процесу аналізу та передбачення добового генерування електроенергії фотоелектростанцією з урахуванням хмарності.

Ключові слова: Модель передбачення, фото електростанція, похибка передбачення, аналіз та передбачення генерації, аналіз даних.

ABSTRACT

Hnatiuk M.Y. Analysis and prediction of daily electricity generation by a photovoltaic power station. Part 2. Analysis of cloud cover influence. Complex master's thesis in specialty 126 - information systems and technologies, educational and professional program - information technology data and image analysis. Vinnytsia: VNTU, 2023. 98 p.

In Ukrainian. Bibliography: 37 titles; figures: 69; tables: 8.

The master's qualification work includes the development of a process for analyzing and predicting the daily generation of electricity by a photovoltaic power station, taking into account cloud cover. Open data with daily aggregations were used. The main regularities have been analyzed. A series of models have been built using machine learning technologies, and the optimal one has been selected. Predictions of electricity generation by the photovoltaic power station have been made, and an analysis of the effectiveness of the optimal model's performance has been carried out.

The illustrative part consists of 11 posters.

In the section of the economic part, the question of the feasibility of developing and implementing a process for analyzing and predicting the daily generation of electricity by a photovoltaic power station, taking into account cloud cover, has been considered.

Key words: Prediction model, photovoltaic station, prediction error, generation analysis, data analysis.

ЗМІСТ

ВСТУП.....	4
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	7
1.1 Аналіз типів сонячних електростанцій та їх конструкцій.....	7
1.2 Проблеми прогнозування генерації електроенергії на ФЕС	13
1.3 Розрахунок прогнозованого виробітку ФЕС з врахуванням впливу хмарності неба у певну годину доби	17
1.4 Огляд і опис методу прогнозування	19
1.5 Вибір оптимальних інформаційних технологій	21
1.6 Огляд та аналіз відомих аналогів.....	24
1.7 Висновки	28
2 ВИБІР ОПТИМАЛЬНИХ НАЛАШТУВАНЬ ДЛЯ РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ	29
2.1 Аналіз можливостей мови програмування Python	29
2.2 Онлайн-платформа Kaggle.....	32
2.3 Підготовка даних	33
2.4 Розвідувальний аналіз	54
2.5 Висновки	63
3 СИСТЕМНИЙ АНАЛІЗ ТА ПЕРЕДБАЧЕННЯ ДОБОВОГО ГЕНЕРУВАННЯ ЕЛЕКТРОЕНЕРГІЇ ФОТОЕЛЕКТРОСТАНЦІЄЮ	64
3.1 Модель об'єкта	64
3.2 Висновки	73
4 ЕКОНОМІЧНА ЧАСТИНА	74
4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки.....	74
4.2 Розрахунок витрат на проведення науково-технічної роботи.....	79
4.2.1 Витрати на оплату праці	79
4.2.2 Відрахування на соціальні заходи.....	81
4.2.3 Сировина та матеріали	82
4.2.4 Розрахунок витрат на комплектуючі	83
4.2.5 Спецустаткування для наукових (експериментальних) робіт	83
4.2.6 Програмне забезпечення та наукових (експериментальних) робіт	83

4.2.7 Амортизація обладнання, програмних засобів та приміщень	83
4.2.8 Паливо та енергія для науково-виробничих цілей	84
4.2.9 Службові відрядження	85
4.2.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації	86
4.2.11 Інші витрати	86
4.2.12 Накладні (загальновиробничі) витрати	86
4.3 Розрахунок економічної ефективності науково-технічної розробки за її можливої комерціалізації потенційним інвестором.....	87
4.4 Висновки	94
ВИСНОВКИ	95
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	97
Додаток А (обов'язковий). Технічне завдання	101
Додаток Б (обов'язковий). Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень	103
Додаток В (довідниковий). Лістинг програми.....	104
Додаток Г (обов'язковий). Ілюстративна частина.....	107

ВСТУП

Актуальність теми. Впровадження відновлюваних джерел енергії (ВДЕ), зокрема фотоелектричних електростанцій, як екологічно чистих джерел, і стимулювання енергетичного ринку спеціальними тарифами для ВДЕ, призвело до швидкого зростання частки ВДЕ в енергетичній системі України.

Найбільш перспективний спосіб отримання, транспортування та використання енергії, яка не завдає шкоди навколишньому середовищу, є сонячна енергетика. Це розділ альтернативної енергетики, що базується на використанні сонячного випромінювання для здобуття енергії у будь-якому форматі. Сонячна енергетика використовує відновлювальні джерела енергії та є екологічно чистою.

За даними Державного агентства з енергоефективності та енергозбереження України [1] очікувані наслідки впровадження сонячної енергетики до 2018 року передбачають збільшення до 1780 ГВт*год, а до 2020 року – до 2420 ГВт*год. Проте зростання рівня стохастичних джерел енергії в енергосистемі супроводжується додатковими ризиками через їх ймовірнісний характер і меншу стабільність, що може призвести до дисбалансу у функціонуванні енергетичної системи і як наслідок втрати балансової стійкості. [2, 3, 4] та якості електроенергії [5]. Відповідно до закону "Про ринок електричної енергії України" [6] для енергогенеруючих компаній, які використовують енергію вітру і сонячного випромінювання та користуються "зеленим тарифом", передбачено встановлення плати за небаланс з доведеними відсотками небалансу, які щорічно зменшуються. З 1 січня 2025 року вона складатиме 100%, а небаланс потужності має знаходитися в межах 5%. Для забезпечення балансової стійкості енергосистеми, необхідно розробити ряд заходів та рекомендацій, які дозволять використовувати регульовані джерела відновлюваної енергії, як частину розподілених віртуальних електростанцій. Це забезпечить стабільність у процесі виробництва електроенергії.

Однією з найбільших проблем є складність передбачення електрогенерації фотоелектростанціє по тій причині, що на сам процес впливає багато фізичних

факторів, враховуючи наведене вище спроби побудови адекватних моделей є дуже актуальними.

Мета і завдання роботи є підвищення точності передбачення добового генерування електроенергії фотоелектростанцією за рахунок використання методів машинного навчання, з урахуванням впливу хмарності.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- провести аналіз закономірностей зміни генерації електроенергії на ФЕС з урахуванням впливу хмарності;
- здійснити вибір оптимальних інформаційних технологій та моделей машинного навчання;
- побудувати моделі зміни генерації електроенергії на ФЕС та вибрати з них оптимальну;
- здійснити передбачення генерації електроенергії на ФЕС з урахуванням впливу хмарності на тестових даних та оцінити ефективність роботи моделі.

Головна задача полягає в тому, щоб за результатами розвідувального аналізу даних побудувати декілька моделей генерації електроенергії на ФЕС та вибрати серед них оптимальну, яка найкраще зможе передбачати генерацію електроенергії ФЕС по доведеному прогнозу погоди (хмарності).

Об'єктом дослідження магістерської кваліфікаційної роботи є процес аналізу даних та передбачення генерації електроенергії на ФЕС з урахуванням впливу хмарності.

Предметом дослідження магістерської кваліфікаційної роботи є методи машинного навчання, які ґрунтуються на сучасних інформаційних технологіях щодо аналізу та передбачення генерації електроенергії ФЕС з урахуванням впливу хмарності.

Новизна одержаних результатів. Дістала подальший розвиток інформаційна технологія передбачення генерації електроенергії фотоелектричною станцією, за рахунок використання методів і моделей машинного навчання та оригінальних ознак на основі хмарності, що дозволяє підвищити точність цього передбачення.

Практичне значення. Практичне значення роботи полягає в тому, що за даними прогнозу погоди, а саме хмарності, інформаційна технологія дозволить передбачити об'єми генерації електроенергії з мінімальною похибкою. Це в свою чергу допоможе уникнути штрафів перед енергоринком та мінімізувати втрати, забезпечуючи стабільну поставку електроенергії.

Апробація результатів магістерської кваліфікаційної роботи. Результати кваліфікаційної роботи доповідались на “Науково-технічній конференції факультету інтелектуальних інформаційних технологій та автоматизації Вінницького національного технічного університету”. (Вінниця, 21–23 червня 2023 р.)

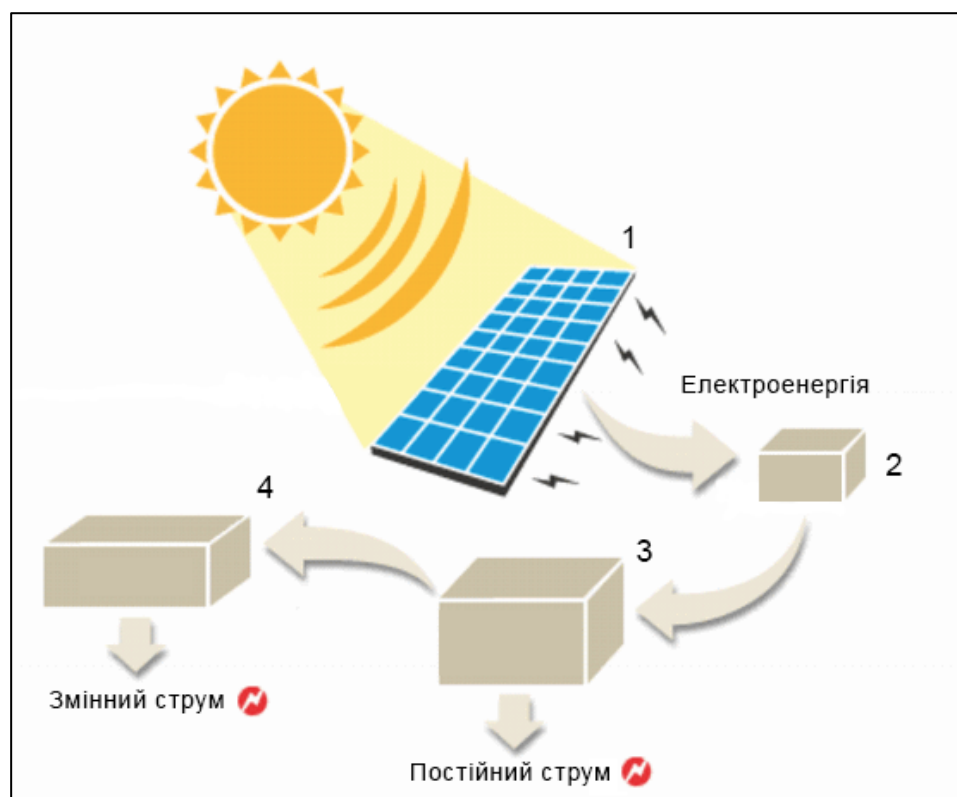
Публікації результатів магістерської кваліфікаційної роботи. В процесі виконання магістерської кваліфікаційної роботи опубліковано тези “Передбачення погодинної генерації електроенергії ФЕС з врахуванням прогнозу хмарності”, у збірнику матеріалів “Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації Вінницького національного технічного університету”. (Вінниця, 21–23 червня 2023 р.) [7].

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Аналіз типів сонячних електростанцій та їх конструкцій

Фотоелектростанція (ФЕС) – це комплекс технічних засобів, призначений для перетворення сонячної енергії в теплову або електричну енергію. На даному етапі існує ряд схем та компоновок фотоелектростанцій, які здатні генерувати електричну енергію з енергії сонця [8].

Головним елементом багатьох фотоелектричних станцій (ФЕС) є сонячні батареї, які включають у себе тонкі шари кремнію чи інших напівпровідникових матеріалів. Ці батареї здатні перетворювати сонячну енергію у постійний електричний струм. Сонячна електростанція складається з таких компонентів: фотоелектричні панелі (які об'єднуються в сонячну батарею), інвертор, контролер, блок акумуляторів і електролічильник, (рис. 1.1).



1 – фотоелектрична панель; 2 – пристрій керування; 3 – акумуляторна батарея; 4 – трьохфазний інвертор

Рисунок 1.1 – Конструкція фотоелектричної сонячної електростанції

Така схема генерації електроенергії є найбільш поширеною. Фотоелектричні панелі перетворюють сонячну енергію на електричну енергію. Акумуляторна батарея слугує для накопичення цієї енергії. Інвертор забезпечує перетворення постійного електричного струму, отриманого від сонячних батарей, у змінний, необхідний для живлення різноманітних електроприладів. Контролер виступає як пульт керування сонячною фотоелектричною системою. Контролер не допускає перевантаження системи або витоку струму в зворотному напрямку у нічний час.

Фотоелектричні перетворювачі (ФЕП) відрізняються своєю надійністю та стабільністю, в свою чергу, основи багатьох світлопоглинальних поверхонь із часом гіршають і втрачають значну частину своєї продуктивності. Сонячні панелі спроможні перетворювати крім прямого, ще і розсіяне сонячне світло. Завдяки невеликій масі, простоті обслуговування та модульному типу конструкції, їх можна застосовувати для створення установок будь-якої потужності.

ККД ФЕС в даний час складає:

- недорогі плівки на аморфному кремнії – 5 %;
- гібридні вироби із двома й більше рідкоземельними елементами – 10-18 %;
- сонячні панелі з моно- або полікристалічного кремнію – 16-19 %;
- багатошарові пристрої з уловлювачами сонячних променів – 40-50 % [8].

Для надійного забезпечення споживача електричною енергією та підвищення ефективності, окрім фотоелектричних панелей, в склад сонячної електростанції входить цілий ряд додаткових елементів, склад яких значно залежить від типу та призначення самої станції. На даний час існують два основних типи сонячних фотоелектричних систем:

- автономні системи, що не пов'язані з енергетичною мережею;
- системи, що генерують та передають електроенергію в електромережу.

Автономна фотоелектрична електростанція, крім сонячних панелей, зазвичай включає акумуляторні батареї (АБ) та контролер заряду/розряду. Для надання електропостачання споживачам, які вимагають стабільної напруги 380В змінного струму, до складу такої станції також необхідно включити інвертор.

Параметри і склад фотоелектричної системи (ФЕС), такі як тип і кількість фотоелектричних панелей, ємність акумуляторних батарей (АБ), потужність інвертора, та вартість системи, залежать від різних факторів: складу електричного навантаження, середньодобового електроспоживання, режиму роботи ФЕС (сезонний або цілорічний), кількості сонячних днів в конкретному регіоні, характеристик сонячної радіації та інші.

Фотопанелі (ФП) генерують енергію тільки протягом світлового дня. Тому їх потужність має бути розрахована так, щоб забезпечити споживачів енергією протягом усього дня. Це призводить до необхідності встановлення більшої потужності ФП та ємності акумуляторів.

Головним недоліком автономних ФЕС є втрати енергії в режимах малих навантажень. У більшості систем ФП вимикаються, коли акумулятори повністю заряджені. Однак можна використовувати надлишкову енергію для підігріву води чи повітря, що трохи підвищує ефективність використання енергії, але не вирішує цілком проблему [9].

Ще однією серйозною проблемою автономних ФЕС є потреба використання акумуляторних батарей, які працюють в циклічному режимі. Більшість кислотно-свинцевих акумуляторів мають обмежену кількість циклів заряду-розряду (1500-2000 циклів), що вимагає їх частої заміни. Використання більш довговічних акумуляторів, таких як нікель-кадмієві або літій-іонні, потребує значних фінансових витрат на побудову енергетичної системи.

Крім того, втрати енергії також відбуваються в акумуляторах. Ефективність процесів заряду-розряду акумуляторів в кращому випадку складає близько 90%, а з часом вона зменшується [8].

Застосування ФЕС у мережі дозволяє уникнути багатьох недоліків автономних систем. По суті, електрична мережа є великим акумулятором з 100% ККД, який може приймати всі надлишки енергії.

Структурна схема мережевої фотоелектричної електростанції зображена на рисунку 1.2. Електрична мережа може бути організована як розподілена або

місцева мережа, яка може бути утворена за допомогою дизельних електростанцій [9].

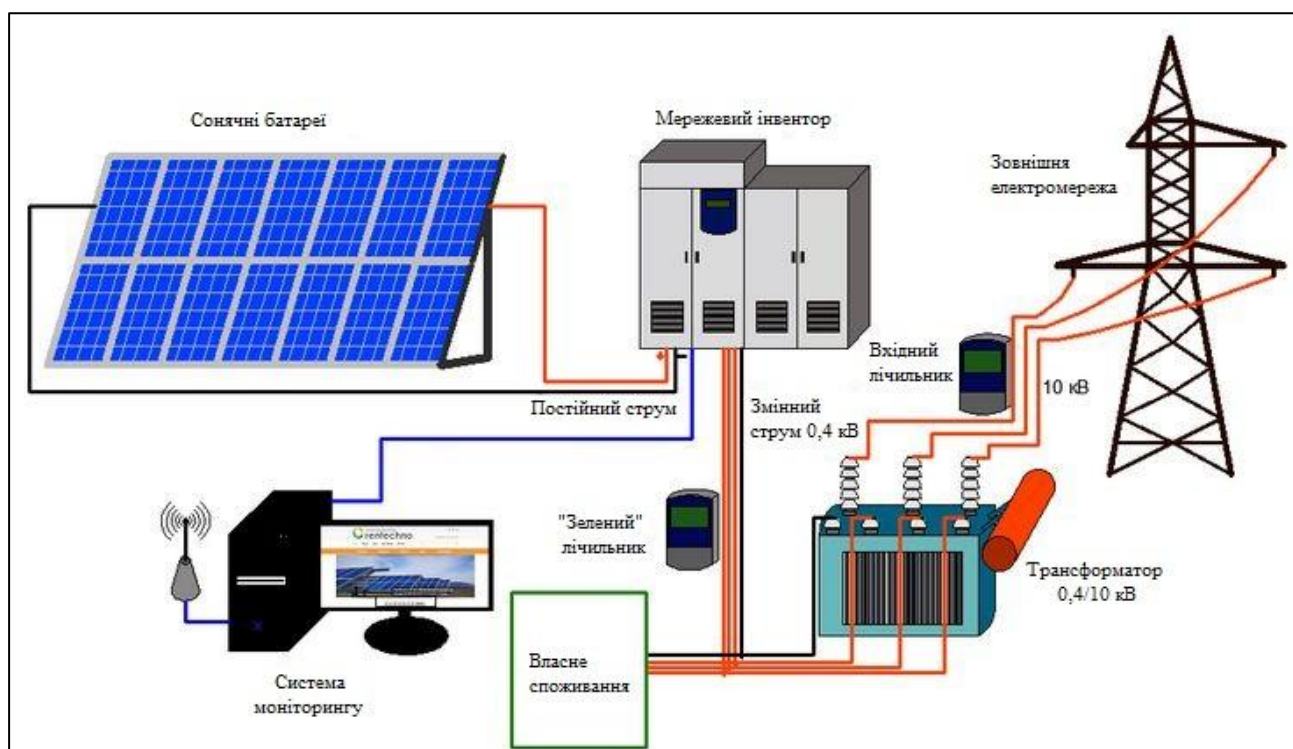


Рисунок 1.2 – Принципова схема мережевої фотоелектричної станції

До складу мережевої сонячної електростанції входять наступні елементи:

- сонячні батареї, що генерують постійний електричний струм під дією сонячної радіації, що потрапляє на поверхню фотоелектричної панелі;
- опорні металоконструкції для розміщення сонячних батарей над поверхнею землі під оптимальним кутом;
- мережеві інвертори, призначені для перетворення постійного струму (DC), що генерується сонячними батареями, в змінний трьохфазний струм (AC);
- лічильники, призначений для обліку генерованої електричної енергії та її продажу за "зеленим" тарифом;
- система моніторингу ФЕС, дозволяє відстежувати роботу фотоелектростанції, забезпечує процес фіксації та збереження отриманих даних від обладнання;

- власні споживачі електроенергії – це та частина електроенергії, що споживається фотоелектростанцією та забезпечує її нормальне функціонування;
- централізована мережа – повітряна або кабельна лінія електропередач, яка зв'язана з енергосистемою і забезпечує електричний зв'язок з фотоелектростанцією.

Поверхню сонячної батареї рекомендується встановлювати перпендикулярно променям сонця. Зазвичай сонячні панелі розміщують на даху або на спеціальних підтримуючих конструкціях. Однак, у зв'язку з цим, важко вести моніторинг стану панелей відносно сонця протягом дня. Тому часто батареї не встановлюються під оптимальним кутом у 90° , який забезпечує найкращу ефективність фотоелектричних систем протягом року незалежно від часу доби [9].

Сонячний трекер – це пристрій, створений для визначення положення сонця та орієнтації несучої конструкції з метою отримання максимальної ефективності сонячної батареї. Концепція трекера полягає в тому, щоб знайти оптимальне положення сонячної батареї по відношенню до Сонця та видавати керуючі команди сервоприводам, для руху конструкцій по одній або двом осям.

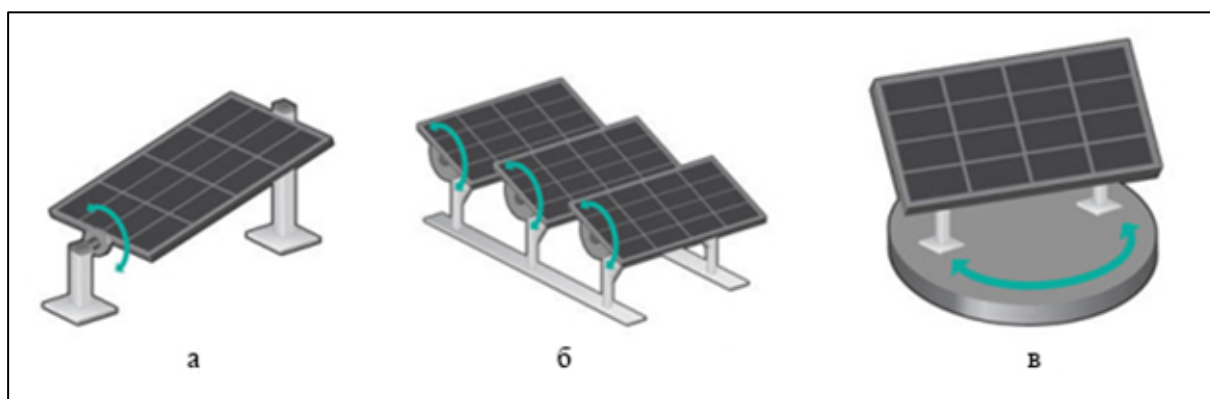
Існують наступні види трекерів: статичні та динамічні. Статичні трекери майже не використовуються через те, що поступаються своїми характеристиками. Динамічні системи найбільш популярні на сьогоднішній день, оскільки забезпечують максимальне використання сонячної енергії. Максимальне виробництво електроенергії відбувається, в положенні коли фотоелектричні панелі розташовані перпендикулярно до сонячних променів.

Фотоелектричні панелі, встановлені на рухомих трекерах, автоматично змінюють своє положення протягом дня, рухаючись за Сонцем, а також автоматично змінюють кут нахилу залежно від пори року.

Існують наступні системи управління трекерами: активні, пасивні та комбіновані. Активна система орієнтується на максимальне сонячне освітлення за допомогою фотоелектричних датчиків. Якщо порівняти їх ефективність із статичними трекерами, ефективність активних систем вища на 49%. Пасивна система орієнтує панелі за сонячним календарем, що дозволяє їм бути ефективнішими, ніж

статичні системи, на 39%. Комбінована система є найбільш надійною серед активних, оскільки переходить у пасивний режим у разі несправностей [10].

Одноосьові сонячні трекери, які мають одну ступінь свободи, характеризуються віссю обертання, орієнтованою на північний меридіан. Вони можуть розширювати свої можливості орієнтації за допомогою додаткових допоміжних пристроїв. Цей тип трекерів може включати горизонтальні, вертикальні, похилі та полярні конфігурації, (рис. 1.3).

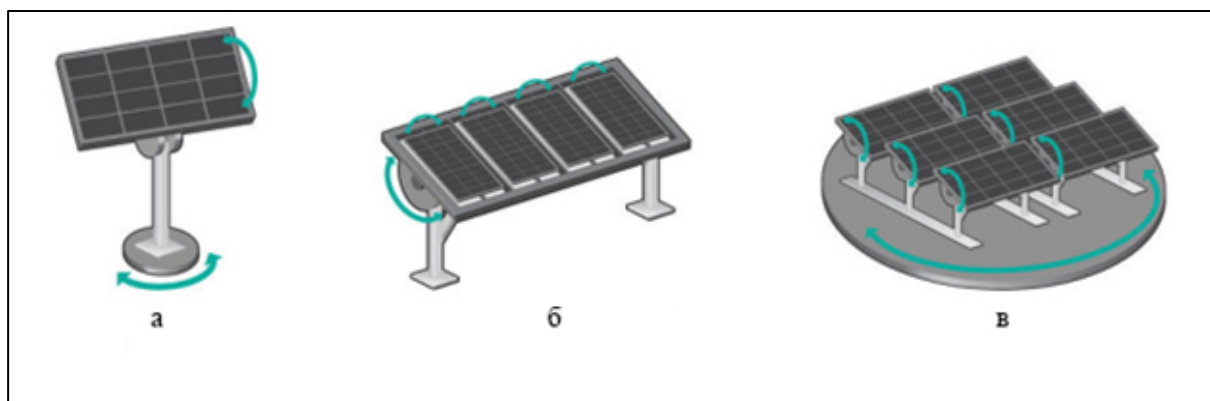


а– похилі; б – горизонтальні; в– вертикальні

Рисунок 1.3 – Види одноосьових трекерів

Двохосьові сонячні трекери мають дві ступені свободи, представлені двома осями обертання. Ці осі функціонують незалежно одна від одної. Одна з осей є основною, а інша використовується як допоміжна. Зазвичай основна вісь є тією, яка фіксується до землі або іншої площини, де буде розташована система.

Двохосьові трекери розподіляються на два типи в залежності від напрямку основної осі щодо землі (рис. 1.4): TTDAT і AADAT. Ці трекери відзначаються автоматичним рухом та вимагають мінімального обслуговування. Їх ефективність перевищує ефективність одноосьових трекерів на близько 30%.



а – класичний; б – рамочний; в – розміщений на обертовій базі

Рисунок 1.4 – Види трекерів з двома осями свободи

Трекер має витримувати сильні вітрові навантаження, опади у вигляді снігу та іншу негоду. Збільшення розміру сонячної батареї призводить до збільшення парусності всієї конструкції, що збільшує навантаження на конструкцію трекера. Щоб вирішити цю проблему, потрібно перерозподілити навантаження на трекер шляхом розширення габаритів всієї конструкції [10].

1.2 Проблеми прогнозування генерації електроенергії на ФЕС

В Україні з 2020 року набули чинності штрафи за небаланси з вироблення електроенергії сонячними електростанціям (рис. 1.5). У зв'язку з цим виникає проблема підвищення точності прогнозування генерації електричної енергії та питання мінімізації штрафів за небаланс між згенерованою кількістю електроенергії та спрогнозованою. Прогнозування стає ключовим інструментом для ефективної інтеграції стохастичних джерел відновлюваної енергії (variable renewable energy, VRE), таких як вітер та сонце, у локальні, регіональні та національні енергетичні системи для досягнення економічної ефективності.



Рисунок 1.5 – Сонячна електростанція «Фрі-Енерджі-Генічеськ»

Проблема узгодження генерації від відновлювальних джерел енергії та традиційних джерел складається з трьох ключових аспектів. По-перше, це справедливе врегулювання прогнозів виробництва від відновлювальних джерел енергії з урахуванням інтересів сторонніх учасників ринку. Другий аспект - точні прогнози споживання енергії. А ось третій аспект включає в себе проблеми короткострокового прогнозування погодних умов "на добу вперед", що фактично вимагає точного прогнозування генерації енергії за годинними інтервалами. Це найскладніша складна частина даної проблеми.

Даний виклик залишається нерозв'язаним надійно та точно ніде у світі. Існує безліч алгоритмів та програмних комплексів, проводяться міжнародні конференції щодо алгоритмів, ведуться дискусії, просуваються нові програмні продукти. Однак не існує загальноприйнятих "стандартів галузі" для прогнозування VRE.

Окрім розв'язанням математичних та обчислювальних труднощів для прогнозування погодних явищ та урахування похибок у межах 5-10%, дуже важливо створити розгалужену інфраструктуру мережі автоматичних станцій для збору метеоданих. Ці станції мають автоматично передавати дані про місцевий стан атмосфери у реальному часі через мережу інтернет. Подальша обробка отриманих даних та розподіл цих даних серед місцевих операторів фотоелектростанцій стають не менш важливими етапами, щоб забезпечити їм необхідну точність та частоту оновлення цієї інформації [11].

Для ілюстрації прикладу складності завдання передбачення генерації, на рисунку 1.6 представлений скріншот системи прогнозування генерації. На станції Xcel Energy розташованій у США використовується 36-годинний горизонт планування, що вважається оптимальним для прогнозування на наступний день. Система автоматичного прогнозування генерації енергії враховує інформацію, що надходить з системи обробки метеоданих. Під час виконання процесу прогнозування на наступний день система враховує фактичну генерацію електроенергії (зелені точки), включаючи миттєву точку прогнозу генерації (суцільна вертикальна лінія), та абсолютну похибку прогнозу на протязі попередніх і наступних семи днів (затінена область). У межах 10% похибки прогнозування ($\pm 5\%$), навіть у найкращому випадку, прогнозується лише наступні 1,5 години, не враховуючи фактичне відхилення електрогенерації від прогнозу погоди.

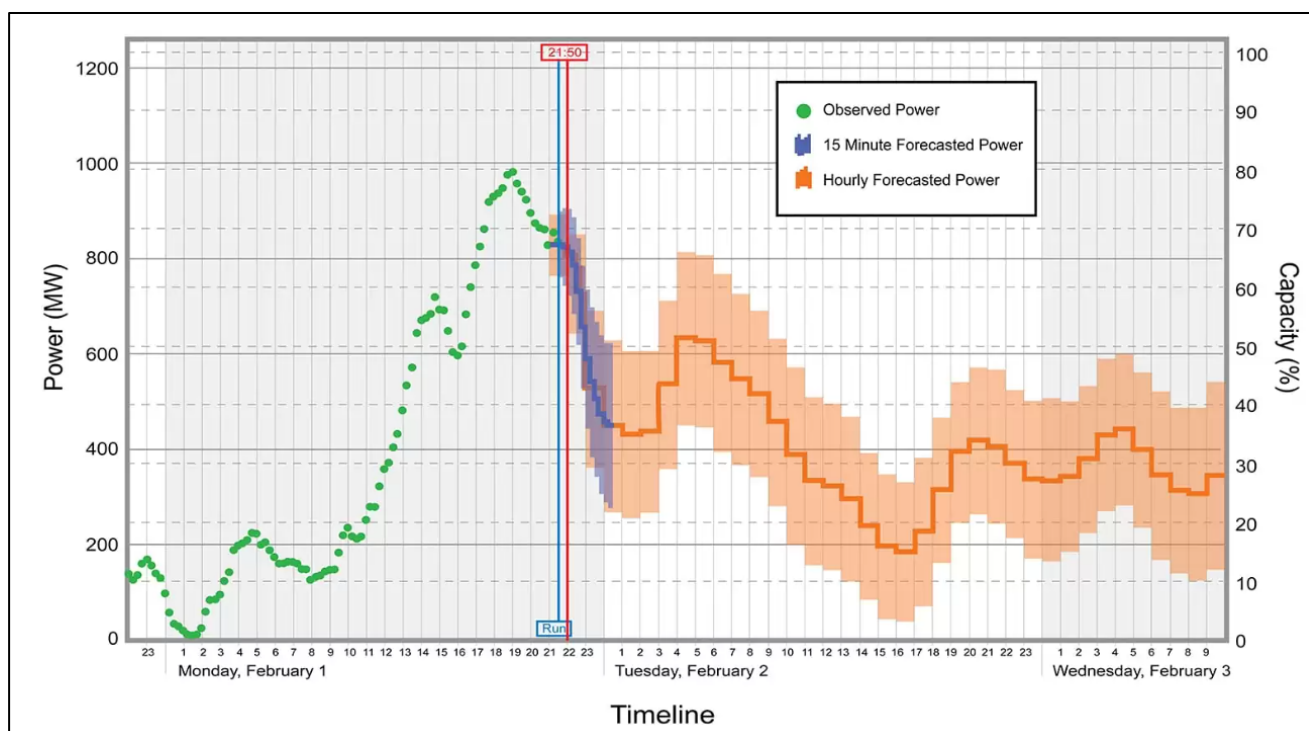


Рисунок 1.6 – Прогнозування генерації станції Xcel Energy у США

Прогнозування генерації електроенергії відновлювальними джерелами енергії (ВДЕ) суттєво впливає на ряд операцій з управління електроенергетичною системою. Це включає планування, диспетчеризацію, балансування в реальному часі, підключення або збільшення резервних потужностей в енергосистемі та видачу команд на запуск або припинення роботи компенсуючих потужностей. Інтеграція прогнозів від локальних операторів, що використовують відновлювальні джерела енергії, дозволяє операторам енергосистеми передбачати швидкі зміни у генерації, щоб ефективно балансувати споживання та заплановану генерацію протягом дня. Це призводить до зменшення залежності від невідновлюваних джерел, що використовують енергію органічного палива та атомних електростанцій, підвищення надійності енергосистеми в цілому та зниження витрат на придбання електроенергії з відновлюваних джерел. Такий підхід сприяє не лише збалансуванню електроенергетичної системи в цілому за потужністю та напругою, а й підвищенню "якості" електроенергії у мережі - зменшенню частотних відхилень та фазових зсувів [11].

1.3 Розрахунок прогнозованого виробітку ФЕС з врахуванням впливу хмарності неба у певну годину доби

Методи прогнозування погоди поділяються на наступні категорії:

– Фізичні методи аналізують дані про погоду, такі як сонячна радіація, хмарність, швидкість та напрямок вітру, враховують рельєф поверхні та перешкоди для створення числових моделей погоди. Наступним кроком, ці числові прогнози використовуються для прогнозування конкретних місцевих погодних умов, що в свою чергу можуть бути перетворені на прогнози щодо виробництва електроенергії від відновлюваних джерел енергії. Такий підхід дозволяє враховувати погодні умови та їх вплив на потенційне виробництво електроенергії з ВДЕ, щоб оптимізувати його використання та забезпечити більш точне планування її виробництва.

- Статистичні методи використовують передусім на історичні дані про виробництво та споживання електроенергії, а також можуть використовувати супутникові знімки або погодні дані для встановлення тенденцій, майже не враховуючи фізичні параметри.

Найкращі результати у прогнозуванні досягаються, коли використовуються обидва підходи. Статистичні методи, що базуються на аналізі історичних даних, та фізичні методи, які враховують фізичні параметри. Комбінування цих підходів дозволяє отримати більш точні та надійні прогнози [11].

Вихідний розрахунок генерації електроенергії фотоелектростанцією враховує умови ясного неба, але фактичні умови можуть суттєво відрізнятись. По тій причині, що абсолютно ясне небо не є типовим для будь-якої години чи дня в році, важливо врахувати вплив хмарності. Навіть легка хмарність може знизити виробництво електричної енергії на рівень до 70%, а сильна хмарність призводить до зниження генерації до 90% [11].

Згідно з джерелом [11], рівень хмарності вимірюється у відсотках. Якщо небо безхмарне, це відповідає нульовій хмарності. У випадку 100% хмарності небо повністю вкрите хмарами. Наприклад, хмарність на рівні 25% означає, що

25% неба прикрито хмарами.

Передбачення хмарності використовує різні числові моделі для прогнозування погоди. Ці моделі базуються на математичних розрахунках атмосфери та океанів, щоб передбачити погоду на основі поточних метеорологічних умов. Загалом, процес створення гідродинамічного прогнозу можна умовно розділити на дві фази:

1) Першу фазу можна назвати підготовчою. Вона розпочинається із збору та архівації даних про погоду в пам'яті комп'ютера метеорологічного центру. Після цього настає фаза декодування зібраних даних, їх контроль якості та формування початкової бази даних. Оскільки розташування метеостанцій неоднакове, дані інтерполюються до вузлів регулярної сітки - точок, рівномірно розташованих в просторі. Багато регіонів на карті погоди, такі як океани або пустелі, з'являються як «білі плями», тому для заповнення цієї недостачі інформації широко використовуються метеорологічні супутники. Результатом цієї підготовчої фази є об'єктивний аналіз метеорологічних даних.

2) Друга фаза, чисельне моделювання процесів в атмосфері, це і є саме гідродинамічний прогноз. Наступним етапом, після об'єктивного аналізу, модель розраховує тенденцію зміни метеорологічних параметрів на короткий період, приблизно до 10 хвилин. Картина погоди, що отримана після розрахунків, є основою для розрахунку подальших тенденцій, і цей процес ітераційно повторюється знову і знову. У більшості метеорологічних центрів максимальний період прогнозу моделі становить 1-2 тижні. Швидкість повторних розрахунків параметрів прогнозу погоди має важливе значення, оскільки прогноз втрачає практичну цінність при занадто тривалій обробці. З метою скорочення машинного часу, багато складних фізичних процесів розраховуються спрощено.

На даний момент існують наступні моделі числового прогнозу погоди:

1) GFS (The Global Forecast System) - це глобальна модель погоди, яка регулярно оновлюється кожні шість годин американською метеорологічною службою. Ця модель складається з чотирьох окремих моделей, що співпрацюють для формування точної карти погодних умов. Комбінування цих моделей

дозволяє отримати більш повну та достовірну інформацію про погоду;

2) ECMWF (European Centre for Medium-Range Weather Forecasts) - це європейська неперервна глобальна модель прогнозу погоди. Вона оперує за концепцією 4D, що дозволяє постійно оновлювати модель по мірі отримання нових супутникових чи інших вхідних даних. Ця підхід дозволяє забезпечувати більш точні та актуальні прогнози, оновлюючи і покращуючи модель у реальному часі з новими вхідними даними;

3) NEMS (meteoblue) - це прогностична модель, що була розроблена у Базельському університеті у Швейцарії, використовуючи моделі NOAA / NCEP (National Centers for Environmental Prediction) як основу. Ця модель включає великомасштабні прогнози, що охоплюють як глобальні, так і локальні території, та істотно покращує прогнозування розвитку хмарності та опадів. Вона є значним вдосконаленням у передбаченні погодних умов, зокрема хмарності та опадів;

4) Модель ICON (Global German Standard) розроблена Німецькою метеорологічною службою (Deutscher Wetterdienst). ICON відома своєю високою точністю, особливо в контексті прогнозування для Європи, іноді вважається ще більш точною, ніж ECMWF. Це пов'язано з кращим покриттям даних, що дозволяє досягти більшої точності в передбаченнях, особливо в межах європейського регіону. [12].

1.4 Огляд і опис методу прогнозування

Машинне навчання – це клас методів штучного інтелекту, де розв'язок завдань формується через навчання системи на основі аналізу багатьох схожих ситуацій чи даних. Для створення цих методів використовують різні математичні інструменти, такі як методи статистики, чисельні методи, оптимізація, теорія ймовірностей, теорія графів, а також техніки обробки цифрових даних. Таке поєднання дозволяє системі вчитися та робити прогнози або приймати рішення на основі набутого досвіду.

Основними задачами машинного навчання є:

- Класифікація: розподілення об'єктів за певними категоріями або класами на основі характеристик. Наприклад, визначення, чи електронний лист є спамом чи не спамом;

- Регресія: прогнозування числового значення на основі залежностей між змінними. Наприклад, передбачення ціни будинку на основі його характеристик.

- Кластеризація: групування об'єктів на основі схожості, без заздалегідь відомих категорій. Наприклад, розділення користувачів за їхніми покупками без попередньо визначених груп;

- Ідентифікація: визначення чи підтвердження ідентичності об'єкта на основі певних параметрів або характеристик;

- Прогнозування: передбачення майбутніх подій або значень на основі даних та виявлення патернів. Наприклад, прогнозування погоди або фінансових ринків.

Задача прогнозування, яка виникла при дослідженні часових рядів, полягає в спробі передбачення майбутніх значень цих рядів на певний часовий проміжок на основі аналізу їх минулих значень [13].

Існують різні методи побудови моделі подібних об'єктів, у тому числі з використанням методі машинного навчання:

- Нейронні мережі: вони базуються на структурі взаємопов'язаних нейронів, що моделюють біологічний мозок. Вони застосовуються для аналізу складних взаємозв'язків між вхідними та вихідними даними, пошуку закономірностей та виявлення структури в даних.

- Древа рішень: цей метод використовує деревоподібну структуру для відображення спостережень та виведення висновків про цільові значення. Вони ефективні для класифікації та прогнозування;

- Баєсові мережі: ймовірнісна графова модель, яка використовує спрямований ациклічний граф для відображення залежностей між випадковими величинами та умовну незалежність між ними. Вони добре підходять для управління невизначеністю та аналізу ймовірностей [14].

У зв'язку з необхідністю знаходження закономірностей у складних багатовимірних випадках, дерева рішень виявляються доречним інструментом для аналізу.

Дерева рішень конструюються за допомогою таких методами:

- бустинг;
- багінг;
- стекінг.

Найбільш ефективним і відомим з них є бустинг. Щодо автоматизації бустингу, існують три найпотужніші бібліотеки:

- XGBoost (Extreme Gradient Boosting): відома своєю ефективністю та швидкістю. Має вбудовані функції для регресії, класифікації та кластеризації;

- LightGBM (Light Gradient Boosting Machine): від Microsoft, відзначається швидкістю та оптимізацією пам'яті, особливо у великих наборах даних;

- AdaBoost: один з перших і відомих алгоритмів бустингу, який використовує слабкі класифікатори та поступово підсилює їх ефективність.

Ці бібліотеки надають інструменти для підвищення ефективності прогнозування шляхом комбінування слабких моделей в сильні. Кожна з них має свої переваги та особливості, що робить їх корисними для різноманітних завдань машинного навчання.

1.5 Вибір оптимальних інформаційних технологій

Передбачення генерації електроенергії ФЕС може бути виконане за допомогою методів машинного навчання, таких як дерева рішень, нейронні мережі та інші моделі. Ці методи дозволяють враховувати вплив різних ознак чи параметрів, таких як метеорологічних умови, час доби, місцезнаходження та інші, для передбачення обсягу електроенергії, яку зможе згенерувати фотоелектрична станція.

Використання нейронних мереж, дерев рішень, та інших моделей у прогнозуванні дозволяє аналізувати складні зв'язки між різними факторами та

робити передбачення стосовно можливих об'ємів генерації електричної енергії у майбутньому.

Аналіз даних складається з наступних етапів:

- збір вхідних даних;
- робота над зібраними даними;
- аналіз даних;
- пояснення отриманих кінцевих результатів.

На етапі обробки даних проводяться різноманітні операції, такі як фільтрування, очищення, перетворення і агрегація, для підготовки даних до подальшого аналізу.

- обчислення характеристик;
- перевірка гіпотез;
- перевірка стохастичності;
- видалення аномальних спостережень;

Під час розвідувального аналізу дані набувають необхідної форми, яка сприяє їхньому кращому розумінню та зручній інтерпретації, а саме в графічному вигляді, таблицях, діаграмах, або схемах. Це сприяє більш глибокому вивченню та розумінню внутрішніх зв'язків у наборі даних [13].

Кореляційний аналіз – це метод в статистиці, який досліджує ступінь взаємозв'язку між двома змінними. Він оцінює, наскільки сильно змінюється одна змінна відносно іншої. Коефіцієнт кореляції приймає значення від -1 до 1: чим ближче значення до 1 або -1, тим сильніше взаємозв'язок між змінними. Значення близьке до 0 вказує на відсутність кореляції.

Кореляційний аналіз має наступні завдання [13]:

- за вибірковими даними оцінюються коефіцієнти кореляції;
- перевіряються на предмет значущості вибіркові коефіцієнти кореляції;
- оцінюється виявлений зв'язок, на скільки він близький до лінійного;
- будується довірчий інтервал для коефіцієнтів кореляції.

Дисперсійний аналіз – це метод, який дозволяє оцінити вплив якісних ознак на результати. Фактори можуть бути неперервними чи дискретними змінними, які поділяються на категорії. Якщо вимірювання на різних рівнях кожного фактора однакове, то аналіз є однорідним, в іншому випадку - нерівномірним.

Задача полягає в дослідженні впливу окремих факторів на мінливість середніх. Він є особливо ефективним при вивченні декількох факторів [14].

Аналіз часових рядів – це дослідження даних, які представляють послідовність значень, зазвичай впорядкованих за часом. Його мета – виявлення закономірностей, трендів, сезонності та інших важливих характеристик, які допомагають у розумінні та прогнозуванні змін у часовому ряді. Для цього використовуються методи статистики, математики та аналізу даних.

Регресійний аналіз – це статистичний метод, що досліджує залежність між змінною, що вивчається (зазвичай позначається як залежна змінна), та однією або більше змінними, які впливають на цю змінну (позначаються як незалежні змінні). Головна мета цього методу полягає в прогнозуванні чи поясненні значень залежної змінної на основі значень однієї чи кількох інших змінних.

Цей метод дає можливість оцінювати природу взаємозв'язку між змінними та будувати модель, що описує цю залежність. Він використовує різні статистичні техніки для аналізу відношень між змінними та встановлення залежностей для передбачення чи пояснення майбутніх значень [15].

Для вирішення поставленої перед нами задачі, по аналізу та передбаченню добового генерування електроенергії фотоелектростанцією з урахуванням хмарності, найдоцільніше буде використати саме регресійні моделі.

Отже, основною метою регресійного аналізу є передбачення конкретних значень, використовуючи наявні дані про різні параметри, такі як рівень хмарності, час року та доби. В нашому випадку, саме ця методика буде використовуватись для передбачення обсягів генерації електроенергії.

1.6 Огляд та аналіз відомих аналогів

Прогноз виробництва електричної енергії фотоелектричними електростанціями значною мірою залежить від аналізу місцевих погодних умов. Вивчаючи кореляції між вихідною потужністю та певними зареєстрованими змінними, можна визначити тенденцію, яка є корисною для прогнозування майбутнього виробництва. Пряме та розсіяне сонячне випромінювання відіграє значну роль у виробництві електроенергії фотоелектричними системами. Однак кореляція між виробництвом електроенергії та іншими факторами, такими як швидкість вітру, температура, час доби та відносна вологість, значно менш виражена через мінливість напрямку та швидкості вітру на різних висотах і відносно стабільні температури протягом короткого періоду часу [16].

Короткострокове прогнозування має вирішальне значення, оскільки фотоелектричні системи часто демонструють значні коливання вихідної потужності протягом коротких періодів часу [17]. Наприклад, перед тим, як сонячну панель затьмарить хмарою, відбувається збільшення виробництва енергії через ефект краю хмари (рис. 1.7). За таких умов вихідна потужність фотоелектричної системи може зрости до 150% від її останнього стабільного значення завдяки спільному впливу прямого сонячного випромінювання та розсіяного випромінювання від сусідніх хмар [18]. Точне передбачення цих явищ дозволяє системі вчасно підготуватися до майбутніх коливань виробництва.

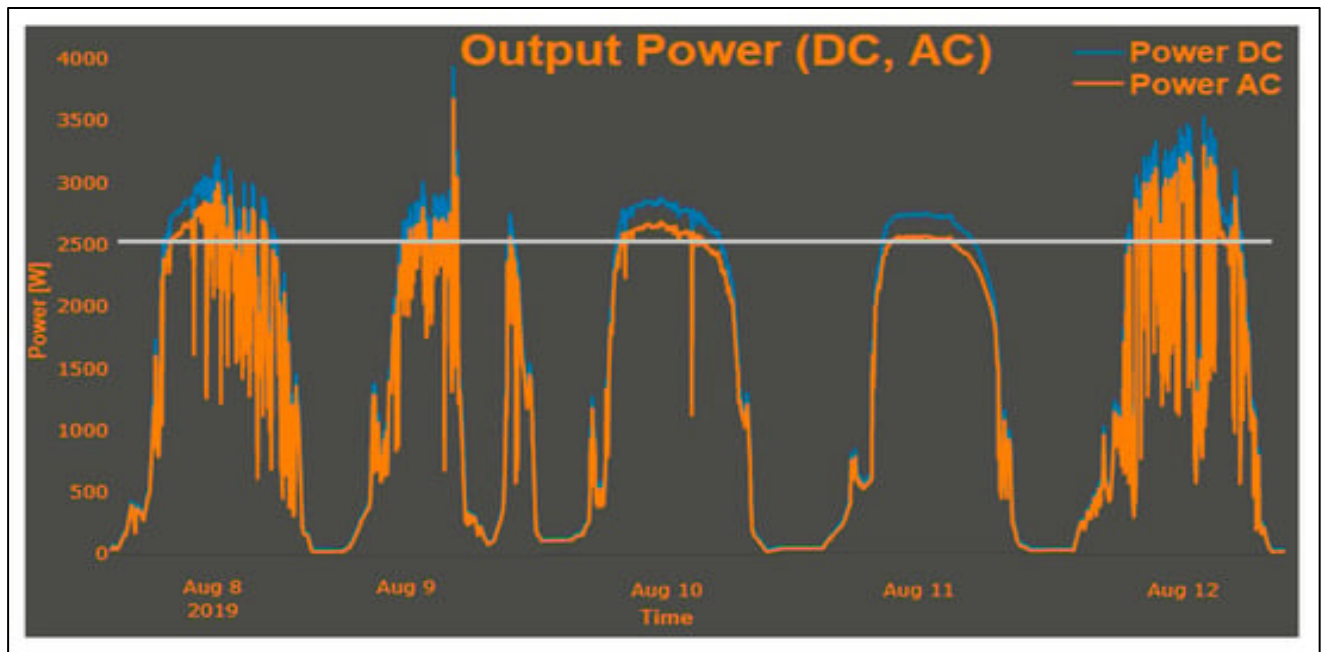


Рисунок 1.7 – Спостереження ефекту краю хмари

Щоб визначити тенденції зміни хмарного покриття, потрібні вхідні дані, які включають поточний і минулий стан хмарного покриття в спостережуваній фотоелектричній системі. Такі елементи, як хмарний покрив, положення, напрямок і швидкість руху хмари, можна отримати шляхом аналізу супутникових і радіолокаційних зображень хмарного покриття [19]. Однак бази даних супутникових знімків погано підходять для короткострокового прогнозування через недостатню часову та просторову роздільну здатність. Для локальних короткострокових прогнозів необхідна база даних із високою просторовою та часовою роздільною здатністю щодо хмарного покриття над об'єктом. Діючим методом створення такої бази даних є серійна фотографія півкулі неба [20].

Оцінка впливу окремих змінних, таких як температура, температура точки роси, відносна вологість, видимість, барометричний тиск, швидкість вітру, хмарний покрив, напрямок вітру та рівень опадів на об'єкті, на вихідну потужність показує низьку кореляцію між окремими змінними і вихідною потужністю. Однак більш значущу кореляцію можна отримати, аналізуючи комбіновані ефекти багатьох змінних [21]. Використання єдиної вхідної змінної, а саме зображення неба, для оцінки поточного та майбутнього результату значно

спрошує систему, підвищуючи її стійкість, легкість впровадження та економічну ефективність. Хоча використання кількох вхідних змінних у поєднанні із зображенням потенційно може призвести до вищої кореляції, цей підхід також ускладнить збір і обробку даних, а також загальний розмір системи для короткострокового прогнозування виробництва фотоелектричної системи.

Крім того, існує кілька методів обробки зображень півкулі неба. Інформація про вектори руху хмар, отримана з послідовності зображень, може підвищити точність системи прогнозування виробництва енергії фотоелектростанцією [22]. Визначення положення хмари в послідовності зображень і порівняння її положення на двох послідовних зображеннях дозволяє визначити вектори руху хмари (рис. 1.8). Беручи до уваги розмір хмари та її напрямок і швидкість руху, можна передбачити можливість майбутнього затінення фотоелектричних панелей хмарою, отже, майбутнє виробництво фотоелектричної енергії в цьому місці [23].

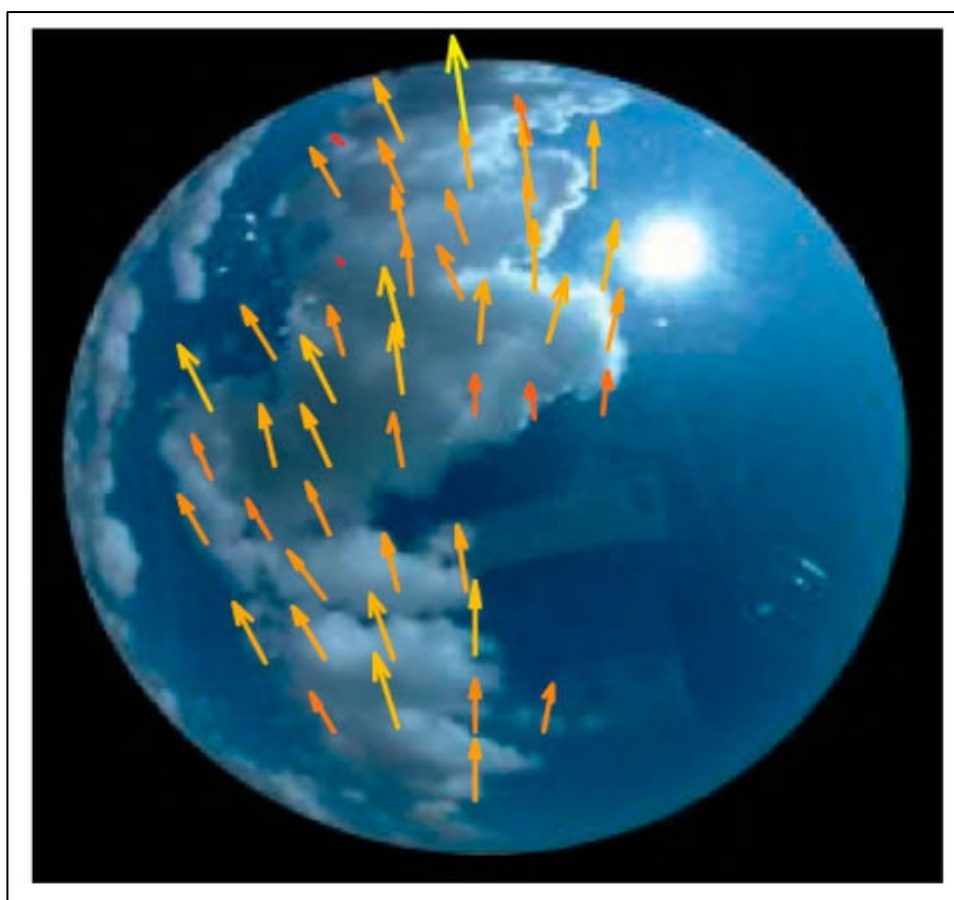


Рисунок 1.8 – Метод вектора руху хмари для короткострокового прогнозування

Використовуючи глибоке навчання можна створити системи, що здатні приймати рішення автоматично. Ця стратегія дозволяє моделям розшифровувати складні шаблони та функції з великих обсягів даних для виконання різноманітних завдань, таких як розпізнавання зображень, аналіз мовлення, мовний переклад тощо [24].

Для короткострокового прогнозування виробництва фотоелектричних електростанцій згортка нейронної мережі (CNN) є особливо вигідною, оскільки вона може декодувати нелінійні зв'язки між входами та виходами моделі. Завдяки здатності мережі розпізнавати візерунки на фотографіях, ці зображення можуть служити вхідними даними [25]. Таким чином, CNN може виявити кореляцію між зображеннями неба та вихідною потужністю фотоелектричної системи. За допомогою LSTM, можна створити модель для короткострокового прогнозування потужності фотоелектричної електростанції [26]. Прогнозуючи майбутні рівні хмарності на основі тенденцій руху хмар і змін умов освітлення, ця модель може точно передбачити майбутнє виробництво фотоелектричних електростанцій.

Для успішного навчання нейронної мережі вирішальною є комплексна високоякісна база даних. Переглядаючи доступні бази даних [27], стає зрозуміло, що бракує повних баз даних зображень неба з високою роздільною здатністю з даними про сонячне опромінення. Низька часова роздільна здатність зібраних фотографій і даних про сонячне випромінювання недостатня для фіксації значних змін вихідної потужності, спричинених затіненням хмар.

База даних, що містить фактичні дані про вихід фотоелектричної електростанції, повинна мати відносно високу часову роздільну здатність, щоб зафіксувати зміни вихідної потужності фотоелектричної системи. Враховуючи високодинамічний характер погодних умов на майданчиках електростанцій через високошвидкісні хмари з відносно невеликими обсягами, необхідні бази даних з високою тимчасовою роздільною здатністю.

Таким чином, використання згорткової нейронної мережі в поєднанні з LSTM є найбільш перспективним для короткострокового передбачення виробництва фотоелектричних електростанцій. Точність моделі, побудованої за

допомогою нейронної мережі, залежить від типу використовуваних вхідних даних, що підкреслює потребу в комплексній і детальній базі даних.

1.7 Висновки

Проведено глибокий аналіз предметної області, розглянуто ряд аналогічних рішень існуючих на даний момент, розглянуто актуальність проблеми передбачення генерування електроенергії ФЕС здійснено формалізацію постановки задачі.

Провівши аналіз інформаційних технологій, прийшли до висновку, що найбільш підходящою для виконання задачі є мова програмування Python. Обрання цієї мови обумовлене її широкою популярністю та великою кількістю бібліотек для аналізу, обробки та прогнозування даних. Також були відібрані конкретні бібліотеки, методи та моделі, які будуть використані для аналізу даних та передбачення генерації електроенергії на сонячних електростанціях.

Для аналізу даних будуть використані бібліотеки NumPy, Pandas і Matplotlib. Щодо передбачення генерації електроенергії на сонячних електростанціях, планується використання моделей, таких як Decision Tree Regressor, Random Forest Regressor і LightGBM Regressor.

2 ВИБІР ОПТИМАЛЬНИХ НАЛАШТУВАНЬ ДЛЯ РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ

2.1 Аналіз можливостей мови програмування Python

З розвитком інновацій у сфері обробки даних, Python з фреймворками даних широко використовується в області машинного навчання та передбачення. Його простота дозволяє виконувати складні обчислення у галузі машинного навчання.

Python - це унікальна, інтерпретована об'єктно-орієнтована мова програмування зі значним наголосом на простоту та потужність. Вона була створена у 1990 році нідерландським інженером-програмістом Гвідо ван Россумом. [28].

Python використовується для різноманітних цілей: від створення ігор та веб-додатків до розробки внутрішніх інструментів для різноманітних завдань. Ця мова також часто використовується в логічних сферах для вивчення та розв'язання практичних питань [28].

Діалекти програмування з часом еволюціонують: інженери додають до них нові можливості та виправляють помилки. Це є способом, за допомогою якого мови адаптуються до нових потреб. Наприклад, код, написаний на Python 2, зазвичай не буде працювати "з коробки" без додаткових змін у Python 3 через різницю у синтаксисі та функціоналі [28].

Однією з можливих недоліків Python є його швидкодія. Python, у порівнянні з іншими мовами, не завжди є найшвидшим виконавцем коду. Код Python спочатку транслюється у внутрішній байт-код, який виконується інтерпретатором Python. У багатьох випадках це може призводити до менш ефективних проектів порівняно з мовами, як от C, які працюють на більш низькому рівні та можуть бути швидшими виконавцями коду.

Так, для більшості додатків пріоритетом є швидкість розробки і просування, а не завжди максимальна швидкодія виконання коду. Python часто використовується через свою простоту та легкість розробки, що дозволяє

створювати програми швидше порівняно з іншими мовами програмування. В більшості випадків для багатьох застосувань швидкість розробки важливіша за найвищу швидкодію виконання.

Python має можливість використовувати модулі, написані на C або C++, які можна використовувати для обробки частин програм, що потребують значного навантаження на процесор. Це дозволяє оптимізувати продуктивність під час виконання важливих операцій.

Python активно розвивається у біологічній системі і може стати ключовим етапом у прикладному машинному навчанні. Однією з головних причин використання Python для аналізу часових рядів є його універсальність: ця мова може бути застосована як для обертання подій, так і для створення.

Python дійсно володіє гнучкістю і можливостями для оптимізації процесів, що зробили його популярним серед програмістів у різних сферах, включаючи аналіз даних, машинне навчання та обробку часових рядів. Python чудово підходить для задач штучного інтелекту в світлі того, що самі обчислення штучного інтелекту важкі для розуміння. Працюючи з Python, дизайнеру не потрібно приділяти надто багато уваги безпосередньо створенню програмного коду, він може зосередитися на вирішенні більш складних та необхідних завдань, пов'язаних із штучним інтелектом.

Порівняно проста лінгвістична структура мови Python полегшує інженерам виконання складних обчислень з відносно невеликим часом виконання. Ще однією перевагою Python є висока якість та доступність документації. У Python існує значна кількість корисних ресурсів, де інженери програмного забезпечення можуть знайти підтримку та наочні пояснення на будь-якому етапі розробки [28].

Python є вигідним у сфері штучного інтелекту через свою адаптивність: інженер може вибирати між різними підходами та сценаріями. Мова допомагає об'єднувати різноманітну інформацію. Крім того, Python є особливо вигідним для тих розробників, які працюють над більшістю коду, використовуючи інтегровані середовища розробки (IDE).

Популярність Python в значній мірі зростає завдяки його простій і чіткій

структурі пунктуації. Через це, велика кількість інженерів, які працюють у галузі машинного навчання, обирають саме Python для своїх проєктів [28].

Перелік основних бібліотек Python:

- Matplotlib. Розробники використовують цю бібліотеку для відображення даних у високоякісній дво- та тривимірній (2D та 3D) графіці. Ця бібліотека поширена під час вирішення наукових завдань. За допомогою цієї бібліотеки, дані можна візуалізувати у вигляді різних діаграм (наприклад, стовпчастих та лінійних). Також можна будувати кілька діаграм відразу, а графіку переносити на будь-які платформи;
- Pandas. Містить оптимізовані та гнучкі структури даних, які можна використовувати для управління даними часових рядів та структурованими даними, такими як таблиці та масиви. Наприклад, Pandas можна використовувати для читання, запису, об'єднання, фільтрації та групування розрахункових даних. Також ця бібліотека широко використовується в науці про дані, аналіз даних та завдання машинного навчання; [29]
- NumPy. Це популярна бібліотека, яка використовується розробниками для простого створення масивів та управління ними, а також управління логічними фігурами та виконання операцій лінійної алгебри. NumPy підтримує інтеграцію з багатьма мовами. Наприклад, C та C++;
- Keras. Це бібліотека глибокої нейронної мережі Python з відмінними функціями обробки даних, візуалізації та багато іншого. Keras підтримує безліч нейронних мереж. Бібліотека має модульну структуру, що забезпечує гнучкість під час написання інноваційних додатків;
- OpenCV-Python. Це бібліотека, яка використовується для обробки зображень під час роботи з машинним зором. Вона містить безліч функцій обробки зображень, таких як одночасне читання та запис зображень, перетворення двовимірного середовища на тривимірне, а також захоплення та аналіз зображень з відео [30, 31].

2.2 Онлайн-платформа Kaggle

Kaggle – це онлайн-платформа спільноти даних, яка об'єднує спеціалістів з аналізу даних, науковців, студентів та ентузіастів з усього світу. Ця платформа створена з метою сприяти розвитку та поширенню знань у сфері машинного навчання, аналізу даних та штучного інтелекту.

Основний фокус Kaggle полягає в організації конкурсів з аналізу даних, де учасники можуть змагатися в розв'язанні складних завдань машинного навчання, використовуючи реальні набори даних. Ці конкурси часто спонсоруються компаніями, які надають дані для аналізу та стимулюють учасників призами чи грошовими винагородами за кращі моделі прогнозування чи розв'язки.

Крім того, Kaggle є великою спільнотою, де користувачі можуть спілкуватися на форумах, обговорювати проекти, ділитися знаннями та досвідом у галузі аналізу даних. Платформа також надає набір інструментів для вивчення та використання різних алгоритмів машинного навчання, статей, блогів та навчальних матеріалів.

Kaggle також пропонує безкоштовні курси з машинного навчання та аналізу даних, що дозволяє початківцям та досвідченим спеціалістам покращити свої навички. Ця платформа виступає не лише як майданчик для змагань, а й як освітній ресурс, що сприяє розвитку та поглибленню знань у сфері аналізу даних та машинного навчання.

Kaggle пропонує користувачам своє власне онлайн-середовище, де можна писати Python/R-скрипти та працювати в Jupyter Notebooks. Робота ведеться у веб-браузері, причому без необхідності встановлення бібліотек та залежностей. Розробки можна поділитися зі спільнотою, також існує можливість оцінювати роботи інших користувачів. Учасники платформи публікують найрізноманітніший контент — від EDA-завдань (Exploratory Data Analysis, тобто детальний розбір) зі змагань до простих методів, які дозволяють оптимізувати власний код.

Вивчення проектів колег допомагає виявити "білі плями" у власних знаннях,

а також зрозуміти, які хард-скили потрібно покращити. Kaggle, таким чином, допомагає відносно швидко удосконалюватися.

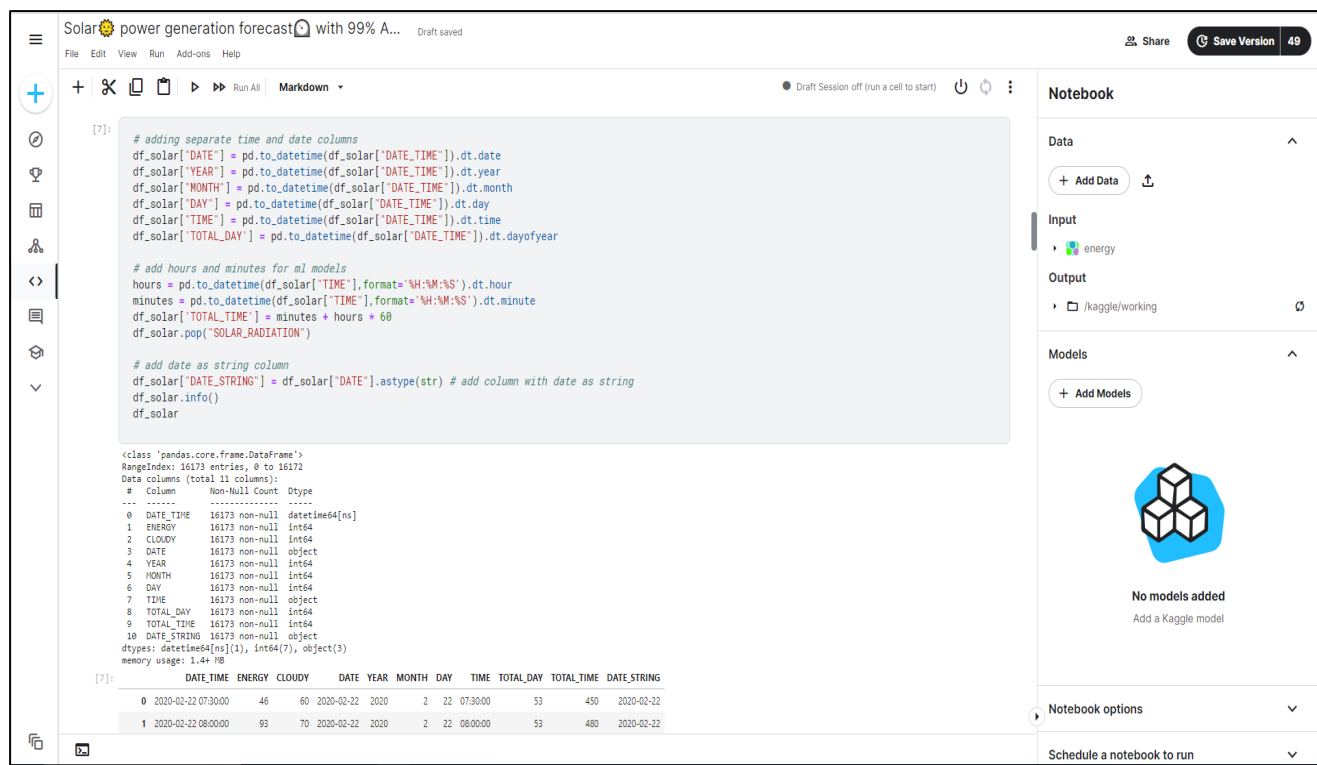


Рисунок 2.1 – Головне вікно онлайн-платформи Kaggle

На рисунку 2.1 зображено головне вікно онлайн-платформи Kaggle. Бокове та горизонтальне меню дозволяє виконувати навігацію в середовищі Kaggle, а також редагувати код, запускати програму та багато іншого функціоналу.

2.3 Підготовка даних

Набір даних моніторингу електрогенерації наданий адміністрацією приватної сонячної електростанції, що розташована на території Вінницької області. Дані від АСКОЕ (автоматична система контролю та обліку електроенергії) фіксуються в системі моніторингу з періодичністю 30 хвилин. Система моніторингу отримує дані лічильників активної енергії та власної метеостанції. Метеостанція фіксує наявні погодні умови, а також дані по сонячній радіації та рівню хмарності.

На першому етапі, необхідно підключити та завантажити пакети маніпуляцій з даними, моделі навчання та інженерні бібліотеки. Фрагмент коду завантаження даних 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 – Завантаження пакетів даних

Підключаємо дані по генерації електроенергії та зафіксованих погодних умов (хмарності та сонячної радіації), (рис. 2.3).

```
# Download data
df_solar = pd.read_csv('../input/energy/DataSetF.csv', sep=';', header=0)
df_solar.head(20)
```

Рисунок 2.3 – Підключення даних

На рисунку 2.4 відображена таблиця даних генерації електроенергії та зафіксованих погодних умов. Де:

«DATE_TIME» – дата та час фіксації показників;

«ENERGY» – об’єм електроенергії, виданий у мережу [кВт*год];

«CLOUDY» – зафіксована хмарність [%];

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

Показник, що найбільше нас цікавить, це «CLOUDY» через те, що навіть невелика хмарність призводить до зниження виробітку ФЕС «ENERGY», отже кількість виробленої електроенергії напряму залежить від хмарності.

	DATE_TIME	ENERGY	CLOUDY	SOLAR_RADIATION
0	22.02.2020 7:30	46	60	49
1	22.02.2020 8:00	93	70	102
2	22.02.2020 8:30	152	60	151
3	22.02.2020 9:00	426	20	467
4	22.02.2020 9:30	406	40	455
...
16168	23.02.2022 15:00	141	70	143
16169	23.02.2022 15:30	77	80	85
16170	23.02.2022 16:00	45	80	48
16171	23.02.2022 16:30	45	70	50
16172	23.02.2022 17:00	7	80	7

16173 rows × 4 columns

Рисунок 2.4 – Таблиця даних

Для початку перетворимо дані в формат «datetime» у форматі “d.m.Y H:M”, (рис. 2.5).

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

Рисунок 2.5 – Перетворення даних в формат «datetime»

Створимо наступні ознаки «YEAR», «DATE», «MONTH», «DAY», «TOTAL_DAY», «TIME», «TOTAL_TIME», (рис. 2.6).

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

# add hours and minutes for ml models
hours = pd.to_datetime(df_solar["TIME"], format='%H:%M:%S').dt.hour
minutes = pd.to_datetime(df_solar["TIME"], format='%H:%M:%S').dt.minute
df_solar['TOTAL_TIME'] = minutes + hours * 60
df_solar.pop("SOLAR_RADIATION")

# add date as string column
df_solar["DATE_STRING"] = df_solar["DATE"].astype(str) # add column with date as string
df_solar.info()
df_solar
```

Рисунок 2.6 – Додавання ознак та перетворення даних

Після додавання додаткових ознак, отримаємо наступний результат, (рис. 2.7, 2.8).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16173 entries, 0 to 16172
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   DATE_TIME       16173 non-null  datetime64[ns]
1   ENERGY         16173 non-null  int64
2   CLOUDY          16173 non-null  int64
3   DATE            16173 non-null  object
4   YEAR            16173 non-null  int64
5   MONTH           16173 non-null  int64
6   DAY             16173 non-null  int64
7   TIME            16173 non-null  object
8   TOTAL_DAY       16173 non-null  int64
9   TOTAL_TIME      16173 non-null  int64
10  DATE_STRING     16173 non-null  object
dtypes: datetime64[ns](1), int64(7), object(3)
memory usage: 1.4+ MB
```

Рисунок 2.7 – Результат перетворення даних

	DATE_TIME	ENERGY	CLOUDY	DATE	YEAR	MONTH	DAY	TIME	TOTAL_DAY	TOTAL_TIME	DATE_STRING
0	2020-02-22 07:30:00	46	60	2020-02-22	2020	2	22	07:30:00	53	450	2020-02-22
1	2020-02-22 08:00:00	93	70	2020-02-22	2020	2	22	08:00:00	53	480	2020-02-22
2	2020-02-22 08:30:00	152	60	2020-02-22	2020	2	22	08:30:00	53	510	2020-02-22
3	2020-02-22 09:00:00	426	20	2020-02-22	2020	2	22	09:00:00	53	540	2020-02-22
4	2020-02-22 09:30:00	406	40	2020-02-22	2020	2	22	09:30:00	53	570	2020-02-22
...
16168	2022-02-23 15:00:00	141	70	2022-02-23	2022	2	23	15:00:00	54	900	2022-02-23
16169	2022-02-23 15:30:00	77	80	2022-02-23	2022	2	23	15:30:00	54	930	2022-02-23
16170	2022-02-23 16:00:00	45	80	2022-02-23	2022	2	23	16:00:00	54	960	2022-02-23
16171	2022-02-23 16:30:00	45	70	2022-02-23	2022	2	23	16:30:00	54	990	2022-02-23
16172	2022-02-23 17:00:00	7	80	2022-02-23	2022	2	23	17:00:00	54	1020	2022-02-23

16173 rows × 11 columns

Рисунок 2.8 – Результат перетворення даних

Проаналізуємо наявні дані та побудуємо графіки для проведення аналізу. На рисунку 2.9 наведено фрагмент коду для побудови добових графіків генерації. Добові графіки генерації зображені на рисунку 2.11-2.13.

```
# Data visualization
solar_dc = df_solar.pivot_table(values='ENERGY', index='TIME', columns='DATE')

def Daywise_plot(data= None, row = None, col = None, title='Energy', 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='blue')

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

Рисунок 2.9 – Побудова добових графіків генерації

На рисунку 2.10 наведено фрагмент коду для побудови добових графіків зміни хмарності. Добові графіки зміни хмарності зображені на рисунках 2.14-2.16.

```

solar_cloudy = df_solar.pivot_table(values='CLOUDY', index='TIME', columns='DATE')

def Daywise_plot(data= None, row = None, col = None, title='Cloudy', 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='blue')
        ax.set_title('{} {}'.format(title, cols[skip_day + i-1]),color='blue')

Daywise_plot(data=solar_cloudy, row=12, col=3, skip_day = 0)

```

Рисунок 2.10 – Побудова добових графіків зміни хмарності

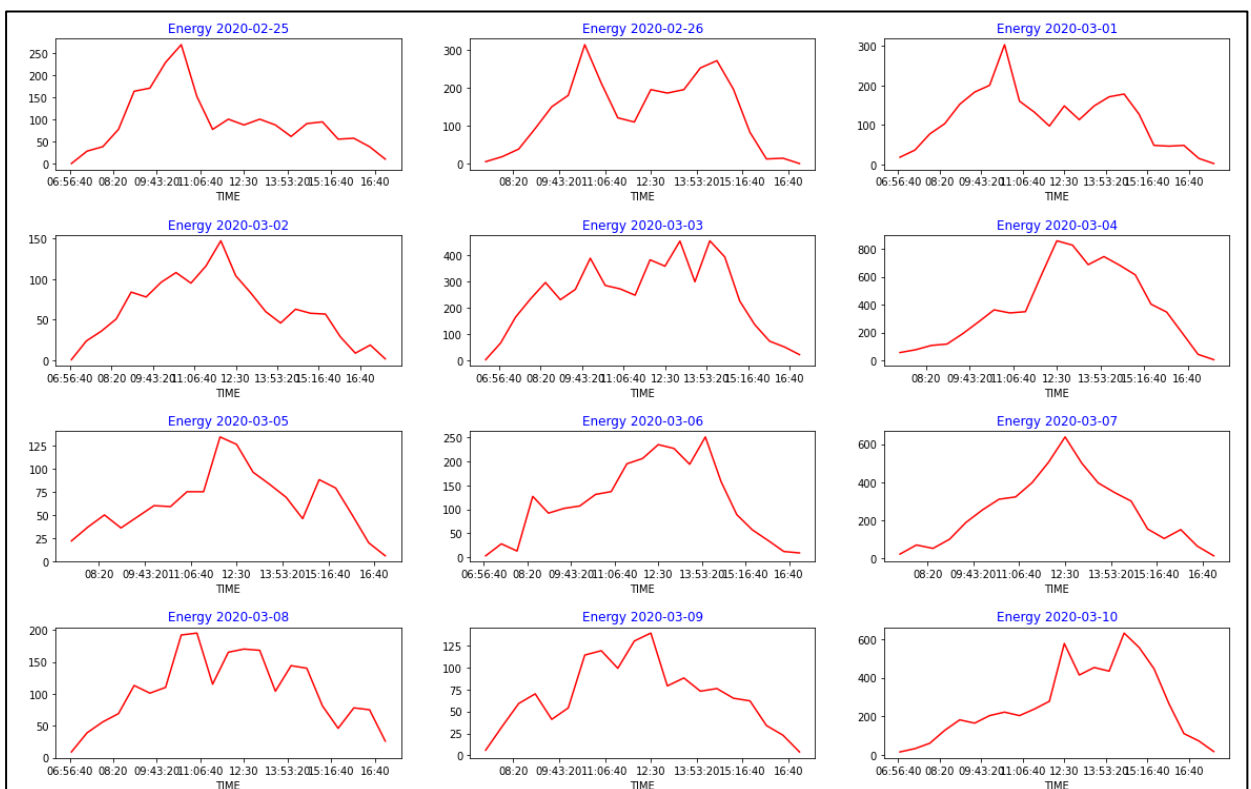


Рисунок 2.11 – Добові графіки генерації за період часу
з 25.02.2020 по 10.03.2020

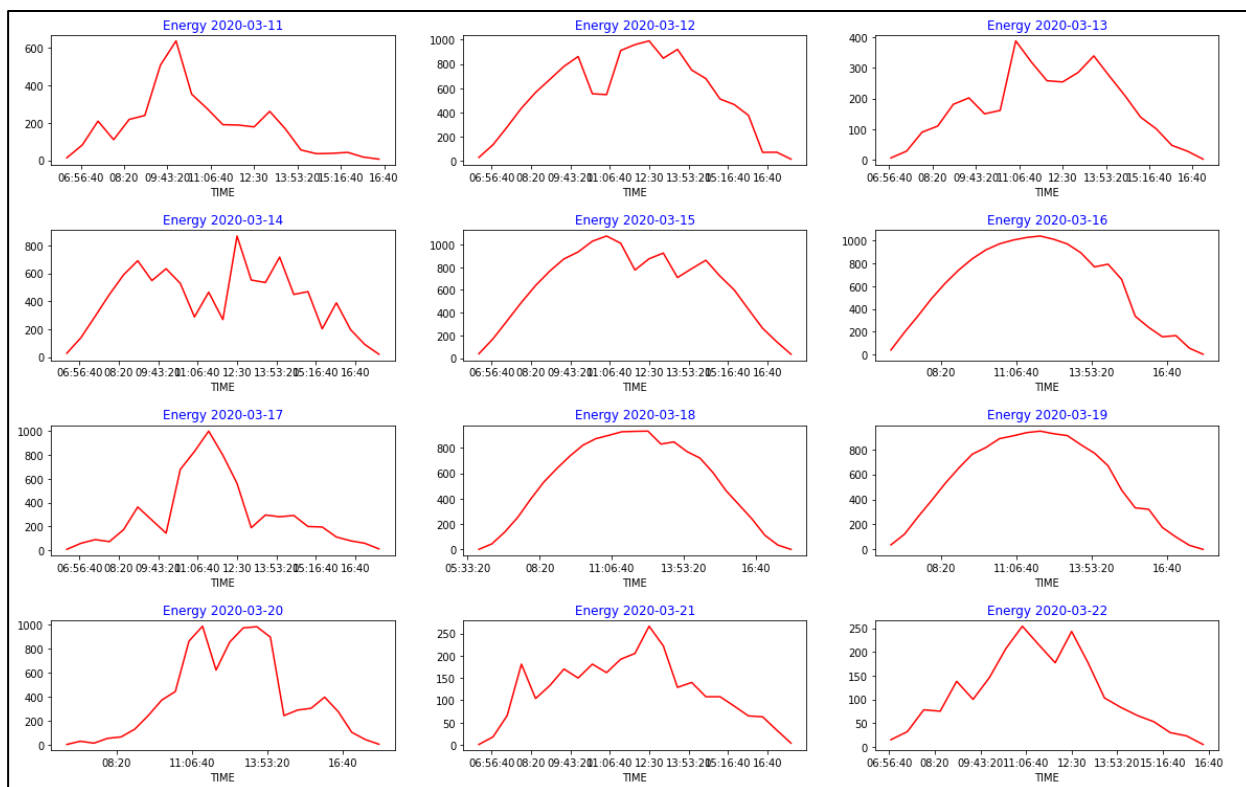


Рисунок 2.12 – Добові графіки генерації за період часу
з 11.03.2020 по 22.03.2020

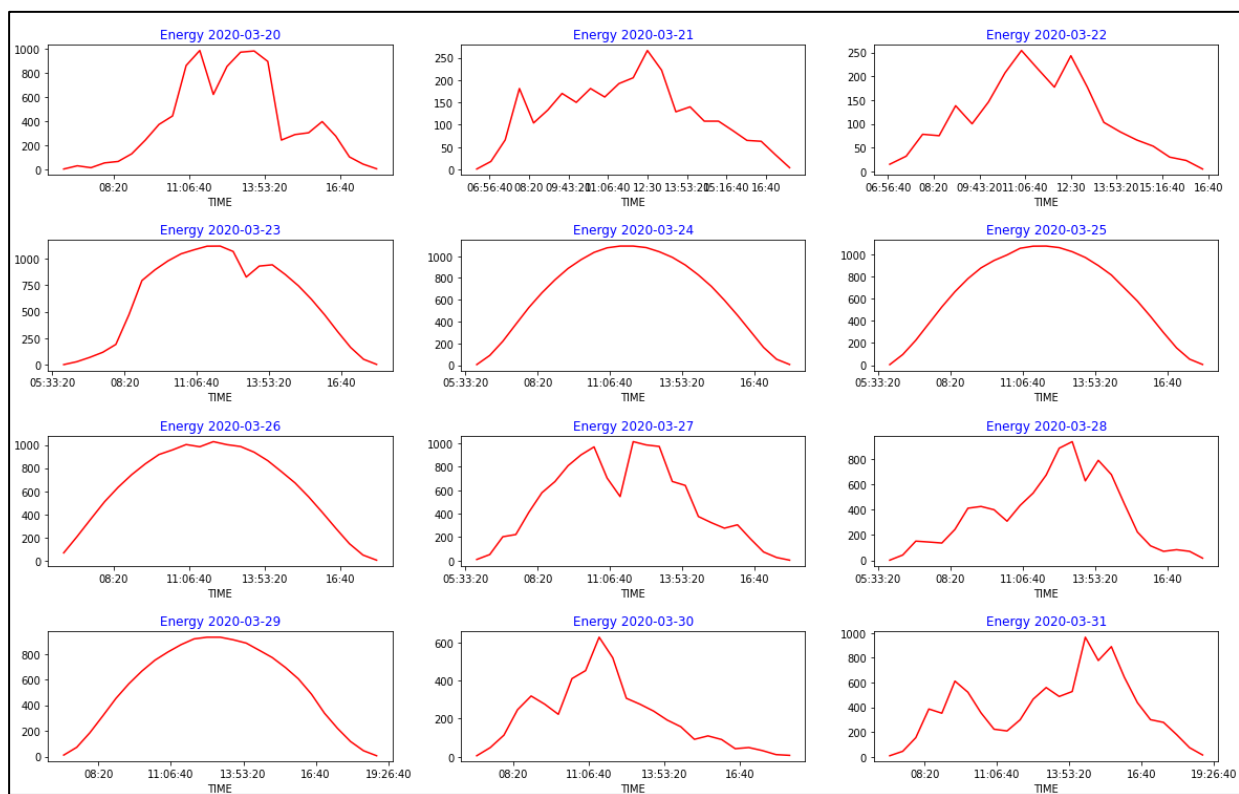


Рисунок 2.13 – Добові графіки генерації за період часу
з 20.03.2020 по 31.03.2020

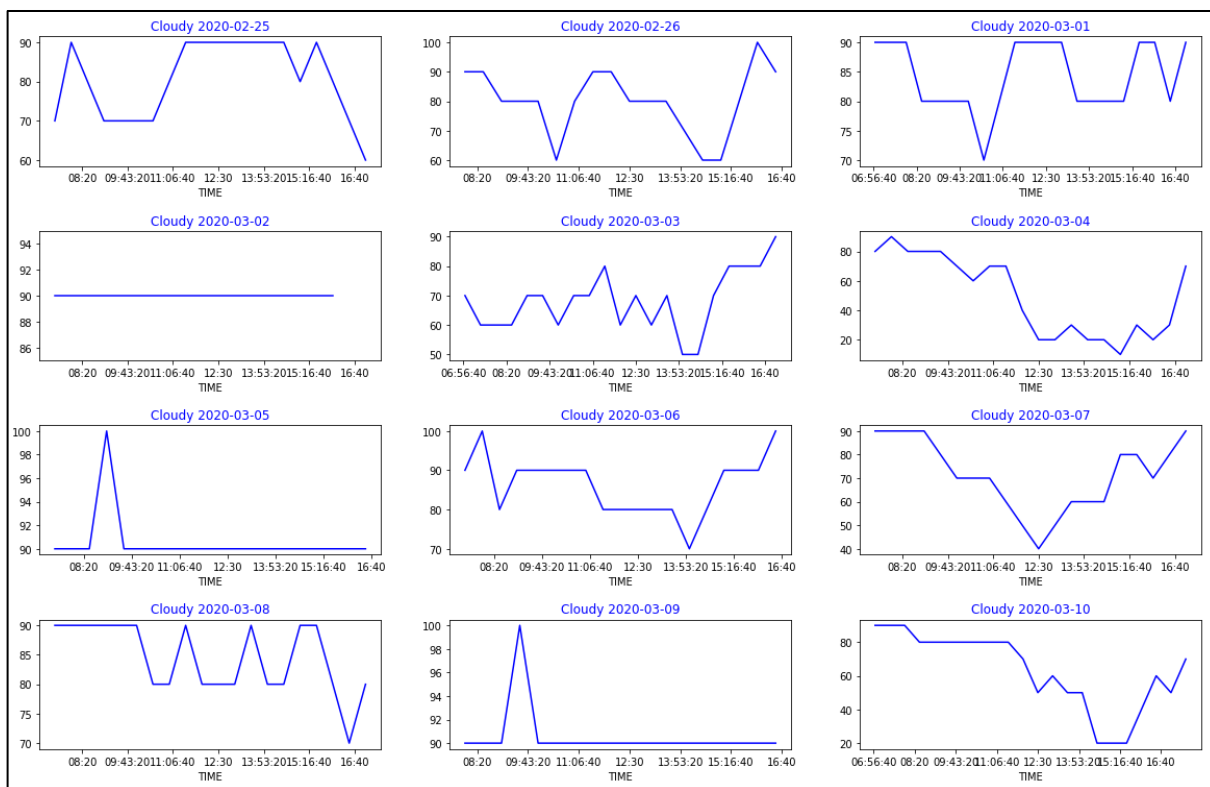


Рисунок 2.14 – Добові графіки зміни хмарності за період часу
з 25.02.2020 по 10.03.2020

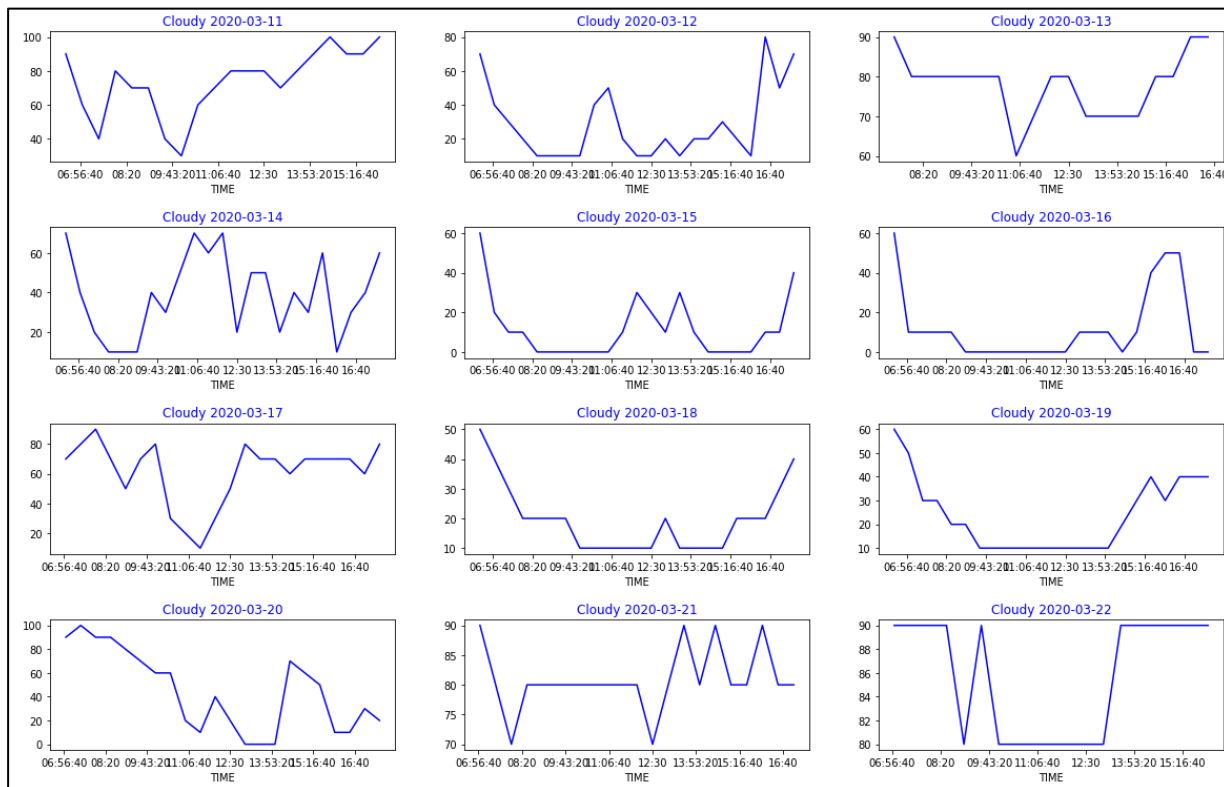


Рисунок 2.15 – Добові графіки зміни хмарності за період часу
з 11.03.2020 по 22.03.2020



Рисунок 2.16 – Добові графіки зміни хмарності за період часу
з 20.03.2020 по 31.03.2020

Як видно з графіків, максимальне значення генерації припадає на час, коли сонце знаходиться у зеніті, а хмарність відсутня, (рис. 2.17).

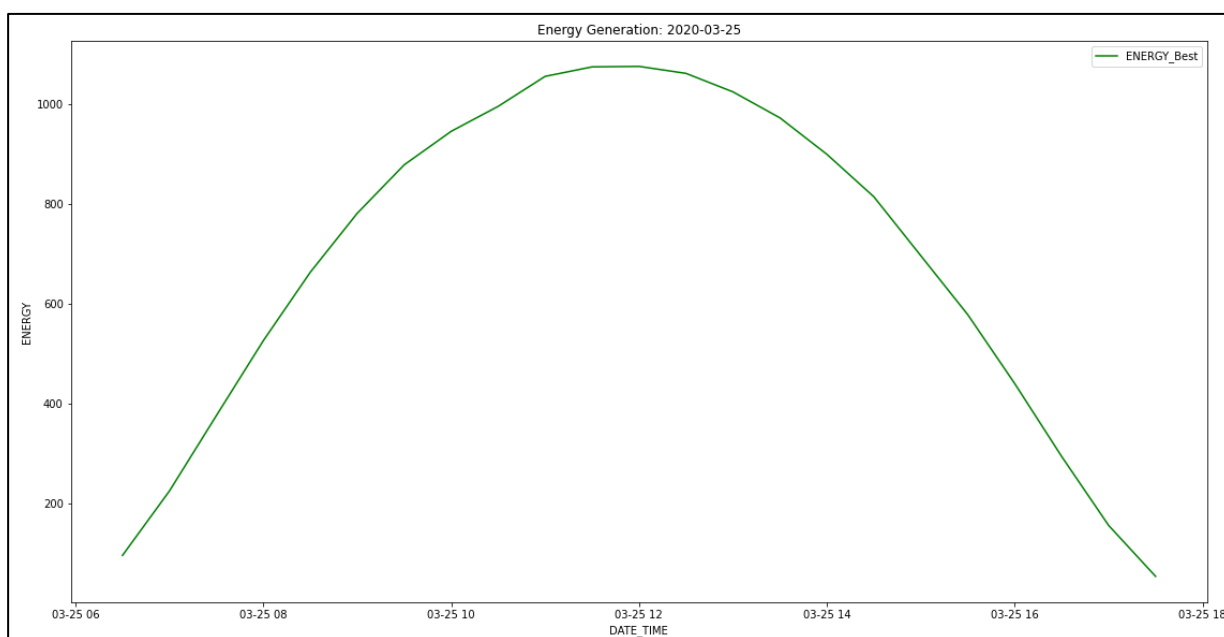


Рисунок 2.17 – Графіки генерації електроенергії при відсутній хмарності

Для того, щоб продовжити аналізувати наявні дані, необхідно дослідити «датасет» на наявність аномалій, які погано впливають на кінцевий результат передбачення, тому необхідно відтворити новий фільтр даних, з метою скорочення різниці між ними. Видаляти аномальні значення (викиди) у датасеті під час машинного навчання важливо з кількох причин:

Вплив на модель: Аномальні значення можуть викривати результати моделі, роблячи їх менш точними. Вони можуть впливати на статистику, яка використовується для побудови моделі, змінюючи середнє значення, дисперсію та інші параметри.

Спотворення результатів: Якщо модель навчання намагається підлаштуватися під аномальні дані, вона може втрачати здатність до генералізації і працювати ефективно на нових даних. Це може призвести до надмірної або неадекватної адаптації моделі.

Ризик виникнення помилок: Аномальні значення можуть бути результатом помилок у вимірюваннях або зборі даних. Вони можуть спотворити аналіз і привести до невірних висновків.

Збереження стабільності моделі: Виключення аномальних даних допомагає моделі бути більш стабільною та надійною, особливо коли вона застосовується до нових, раніше невідомих даних.

Покращення результатів: У випадку, коли аномальні дані впливають на результати, їх видалення може покращити прогностичні можливості моделі, забезпечуючи більш точні прогнози.

Звісно, видалення аномальних значень потребує обережності і обґрунтованих рішень. У деяких випадках видалення може бути не потрібним, особливо якщо аномалії мають важливий контекст або пояснення. Однак у більшості випадків це важлива частина попередньої обробки даних для досягнення кращих результатів у машинному навчанні.

Побудуємо погодинний графік генерації електроенергії, згрупувавши дані по «TIME», де значення енергії «ENERGY» вибираються за максимальним значенням. Фрагмент коду та результат наведено на рисунку 2.18.

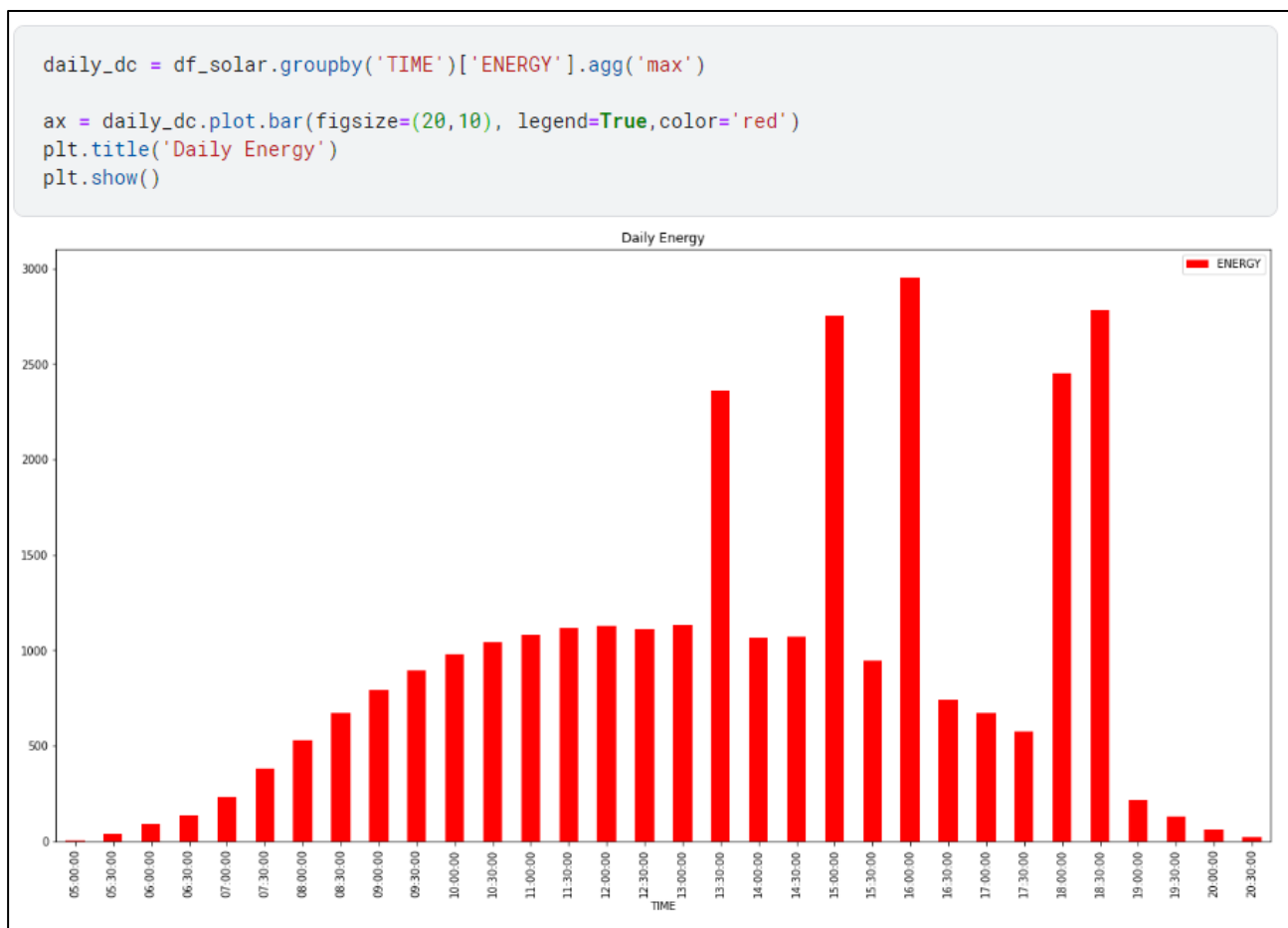


Рисунок 2.18 – Фрагмент коду пошуку аномалій за максимальним значенням та результат

Виходячи з того, що встановлена потужність ФЕС 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 [11]. Значення генерації менші за 10 кВт*год, також вважаємо аномальними та будемо видаляти. На рисунку 2.19 наведено фрагмент коду пошуку дат з аномально високим значенням електрогенерації.

```
df_anomal = df_solar.loc[(df_solar["ENERGY"] > 1200)]
df_anomal
```

	DATE_TIME	ENERGY	CLOUDY	DATE	YEAR	MONTH	DAY	TIME	TOTAL_DAY	TOTAL_TIME	DATE_STRING
4719	2020-08-14 18:00:00	2452	0	2020-08-14	2020	8	14	18:00:00	227	1080	2020-08-14
4720	2020-08-14 18:30:00	2784	0	2020-08-14	2020	8	14	18:30:00	227	1110	2020-08-14
8384	2021-03-04 15:00:00	2752	40	2021-03-04	2021	3	4	15:00:00	63	900	2021-03-04
8386	2021-03-04 16:00:00	2517	60	2021-03-04	2021	3	4	16:00:00	63	960	2021-03-04
12545	2021-08-06 16:00:00	2954	30	2021-08-06	2021	8	6	16:00:00	218	960	2021-08-06
14753	2021-11-20 13:30:00	2358	50	2021-11-20	2021	11	20	13:30:00	324	810	2021-11-20

Рисунок 2.19 – Фрагмент коду пошуку дати з аномально високим значенням електрогенерації

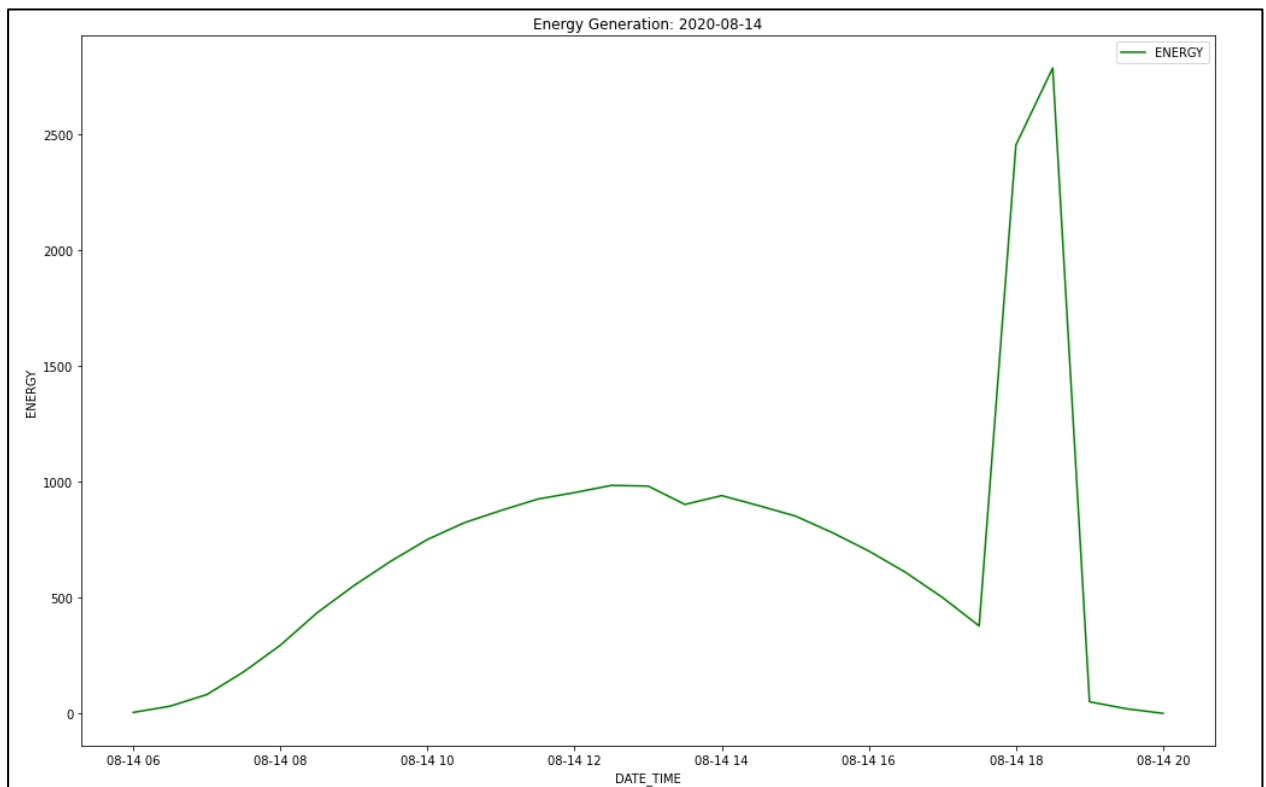


Рисунок 2.20 – Добовий графік електрогенерації за 14.08.2020 з присутньою аномалією

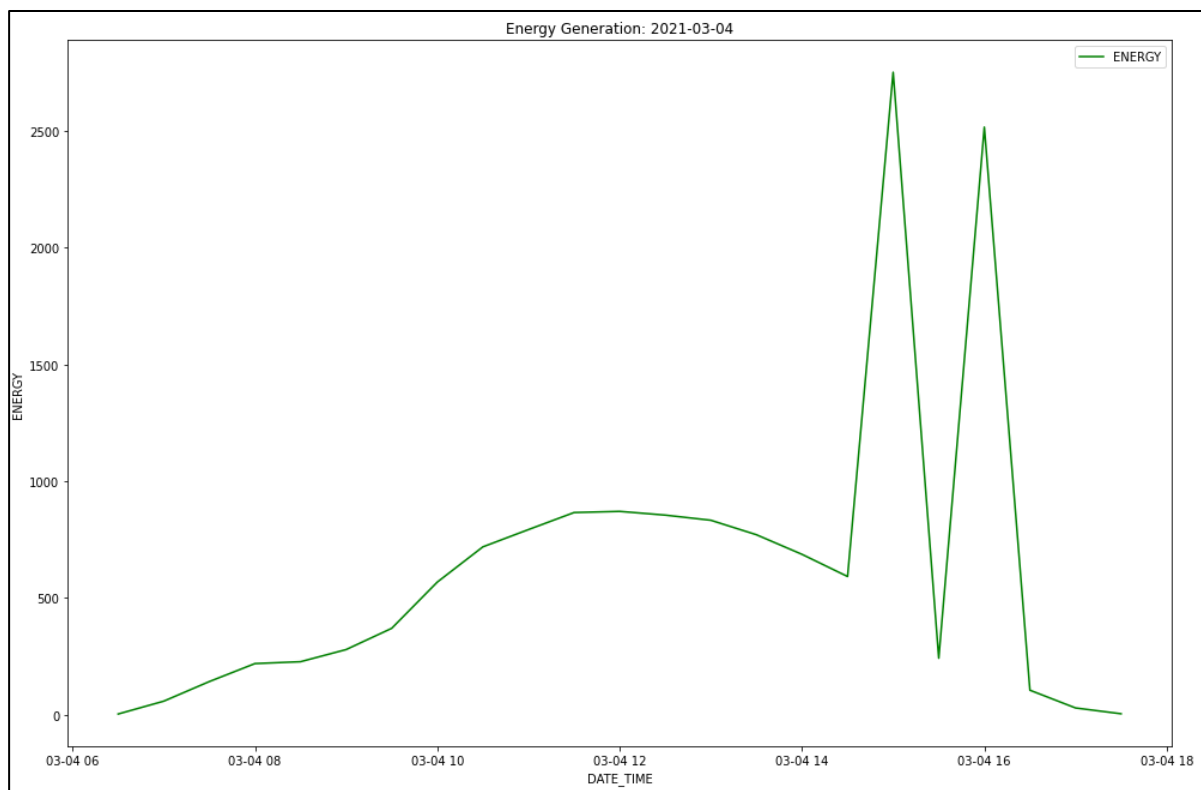


Рисунок 2.21 – Добовий графік електрогенерації за 04.03.2021
з присутньою аномалією

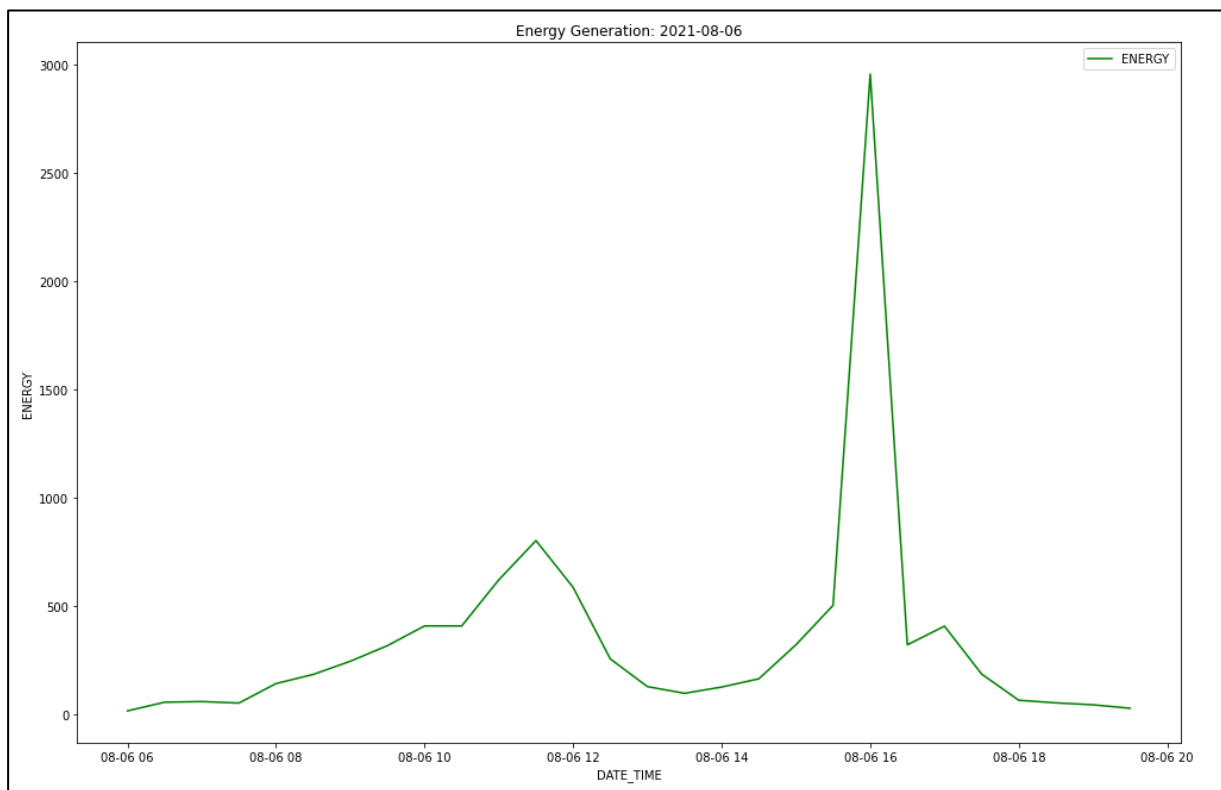


Рисунок 2.22 – Добовий графік електрогенерації за 06.08.2021
з присутньою аномалією

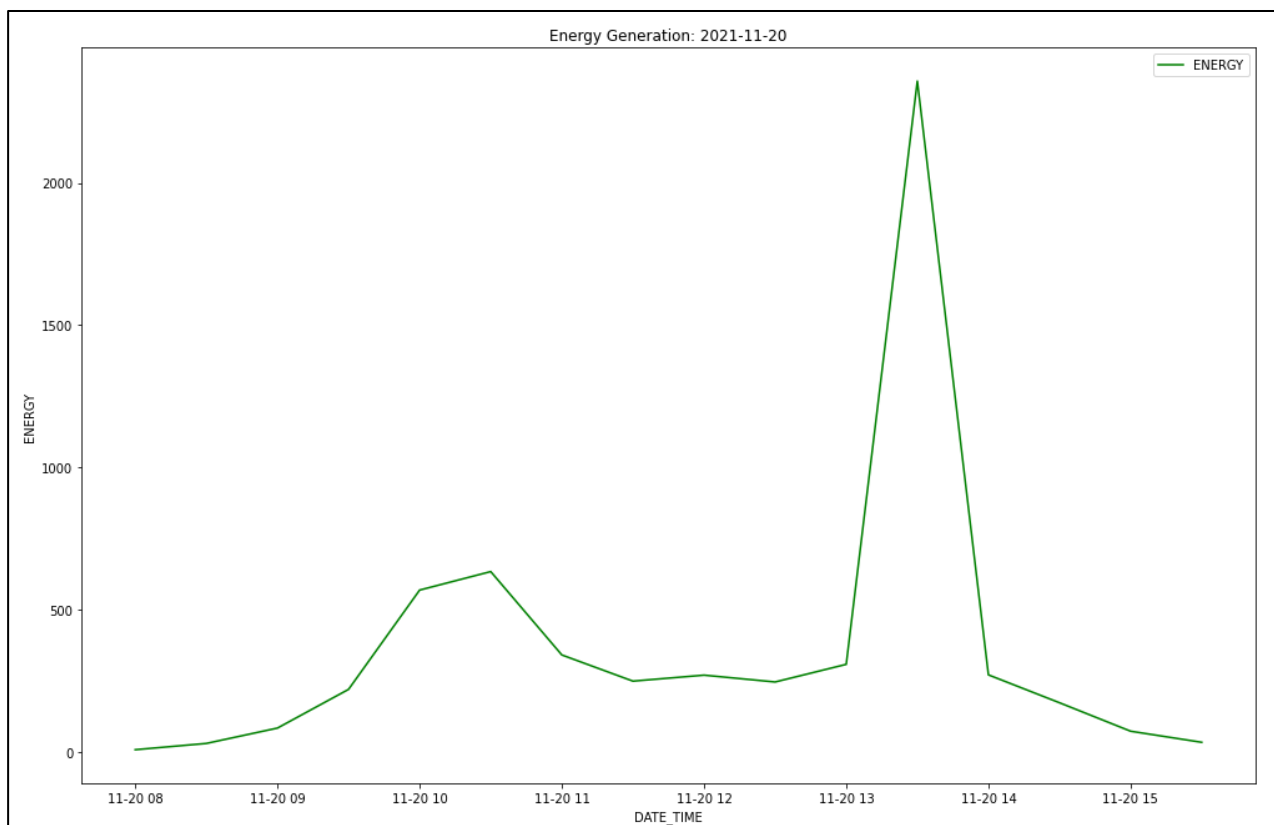


Рисунок 2.23 – Добовий графік електрогенерації за 20.11.2021
з присутньою аномалією

Зафіксовані аномалії, обумовлені несправністю в системі фіксації поточних значень потужності. Як було зазначено вище, значення потужності не може перевищувати 1200 кВт*год, тому наявні аномалії необхідно видалити з датасету, для підвищення точності розрахункової моделі (рис. 2.20 – 2.23).

Значення електрогенерації, що не перевищують 10 кВт*год також є аномальними, це обумовлено похибкою вимірювальних приладів в момент сходу і заходу Сонця. На рисунку 2.24 наведено фрагмент коду пошуку дат з аномально низьким значенням електрогенерації в світлу пору доби.

```
df_anomal = df_solar.loc[((df_solar["ENERGY"] < 10)
                        & (df_solar["CLOUDY"] < 90)
                        & (df_solar["TOTAL_TIME"] > 540)
                        & (df_solar["TOTAL_TIME"] < 960))]
```

df_anomal

	DATE_TIME	ENERGY	CLOUDY	DATE	YEAR	MONTH	DAY	TIME	TOTAL_DAY	TOTAL_TIME	DATE_STRING
51	2020-02-24 15:00:00	9	70	2020-02-24	2020	2	24	15:00:00	55	900	2020-02-24
3588	2020-07-07 10:00:00	8	80	2020-07-07	2020	7	7	10:00:00	189	600	2020-07-07
5914	2020-10-01 15:00:00	6	60	2020-10-01	2020	10	1	15:00:00	275	900	2020-10-01
5974	2020-10-04 14:30:00	2	60	2020-10-04	2020	10	4	14:30:00	278	870	2020-10-04
6012	2020-10-06 14:30:00	3	80	2020-10-06	2020	10	6	14:30:00	280	870	2020-10-06
...
15465	2022-01-13 15:30:00	8	80	2022-01-13	2022	1	13	15:30:00	13	930	2022-01-13
15493	2022-01-15 15:30:00	6	80	2022-01-15	2022	1	15	15:30:00	15	930	2022-01-15
15523	2022-01-17 15:00:00	7	80	2022-01-17	2022	1	17	15:00:00	17	900	2022-01-17
15569	2022-01-20 15:00:00	6	20	2022-01-20	2022	1	20	15:00:00	20	900	2022-01-20
15584	2022-01-21 15:30:00	9	80	2022-01-21	2022	1	21	15:30:00	21	930	2022-01-21

108 rows × 11 columns

Рисунок 2.24 – Фрагмент коду пошуку дати з аномально низьким значенням електрогенерації

На рисунку 2.25 відображено графік з аномалією, обумовленою аварійною ситуацією (відключення фотоелектростанції від мережі), при цьому потужність впала до нульових значень в світлу пору доби.

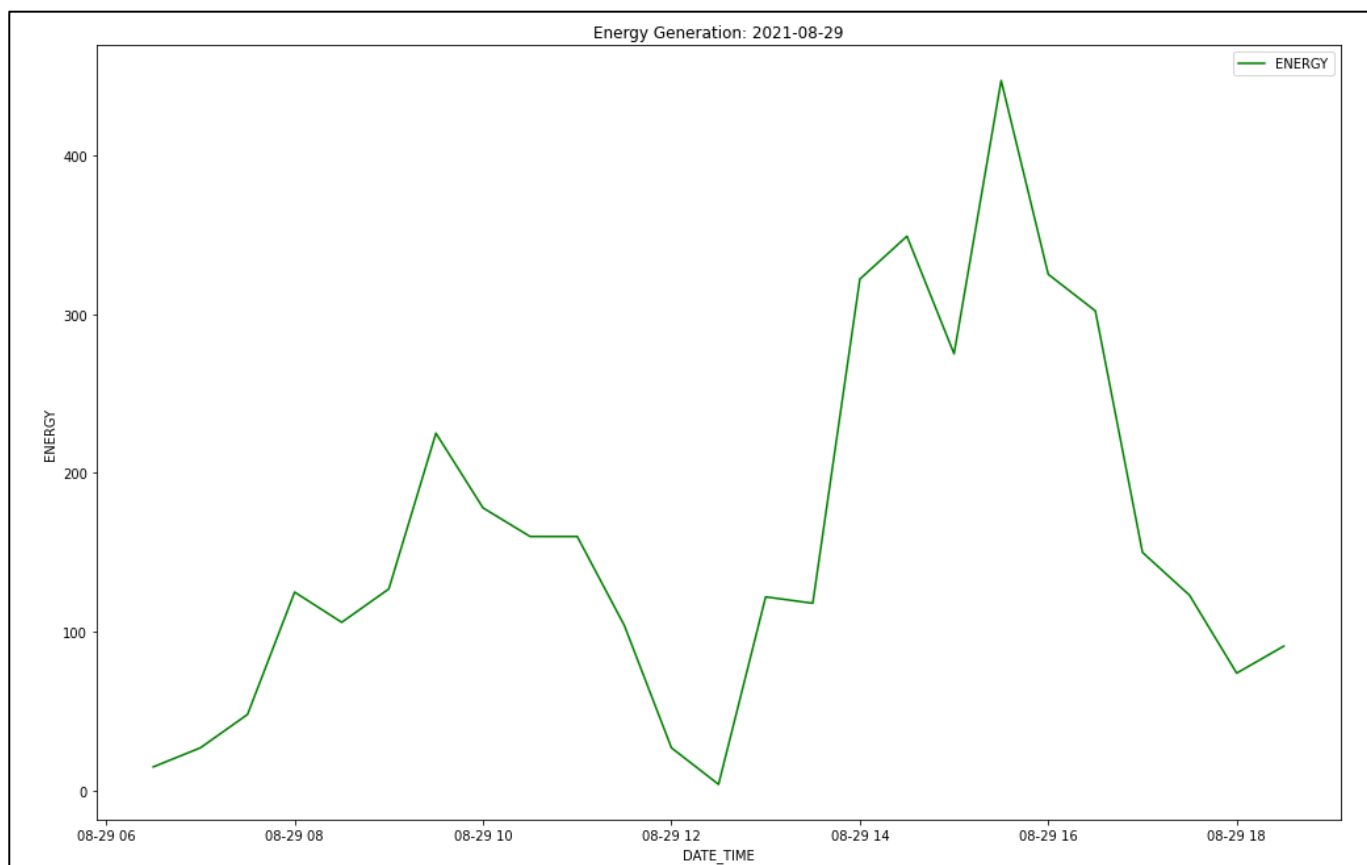


Рисунок 2.25 – Добовий графік електрогенерації за 29.08.2021
з присутньою аномалією

Також одною з причин занадто низьких показників електрогенерації можуть бути опади, у вигляді снігу, на поверхні фотоелектричних панелей. На рисунках 2.26 – 2.28 відображено графік електрогенерації з зазначеною аномалією.

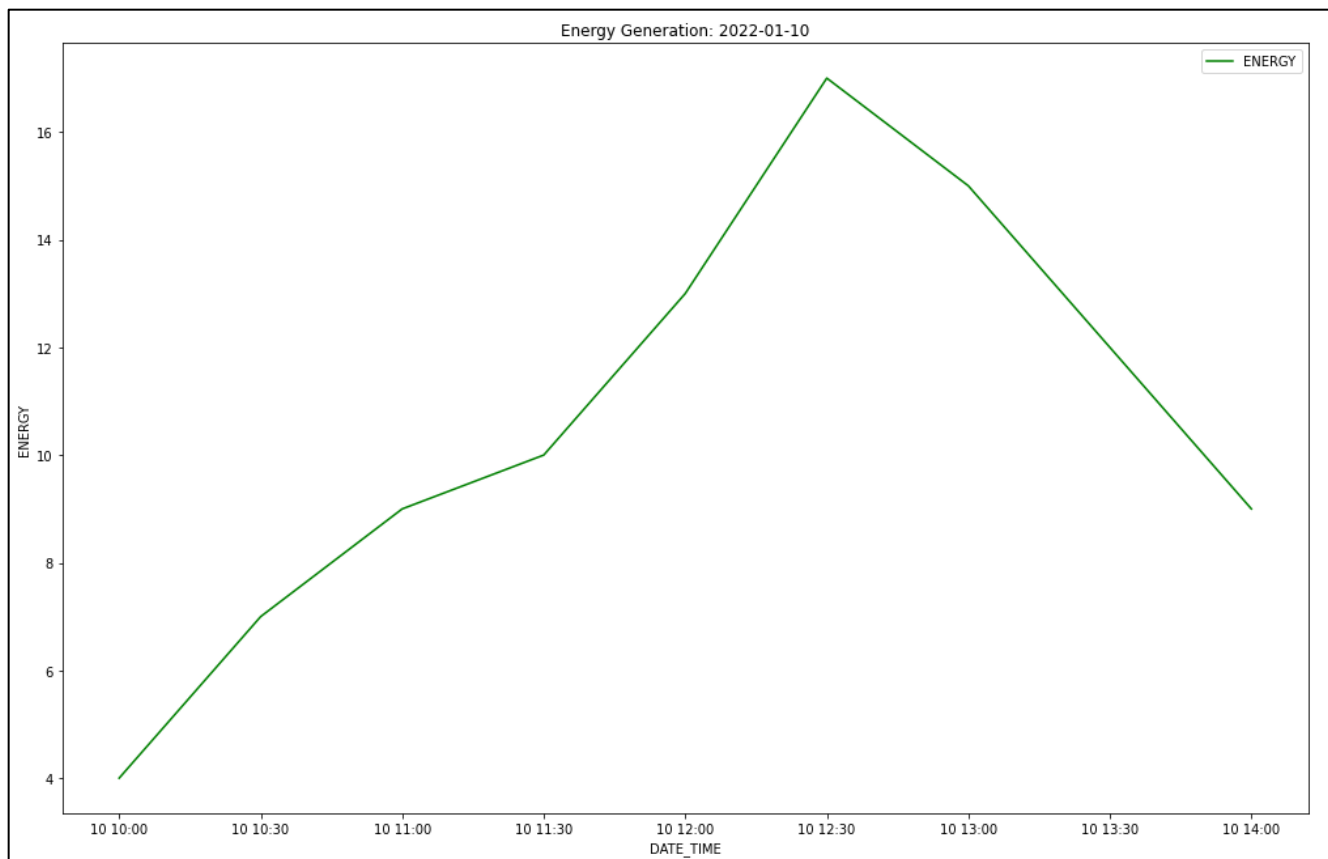


Рисунок 2.26 – Добовий графік електрогенерації за 10.01.2021
з присутньою аномалією

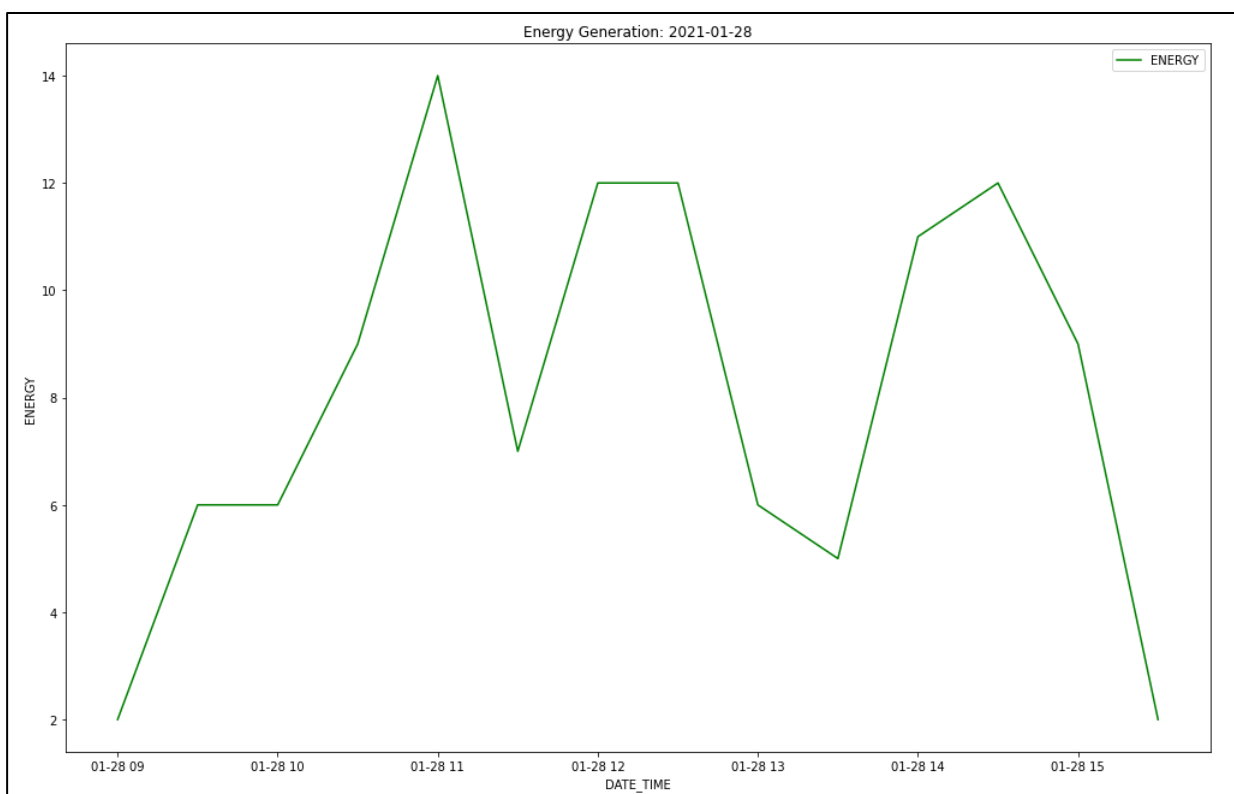


Рисунок 2.27 – Добовий графік електрогенерації за 28.01.2021
з присутньою аномалією

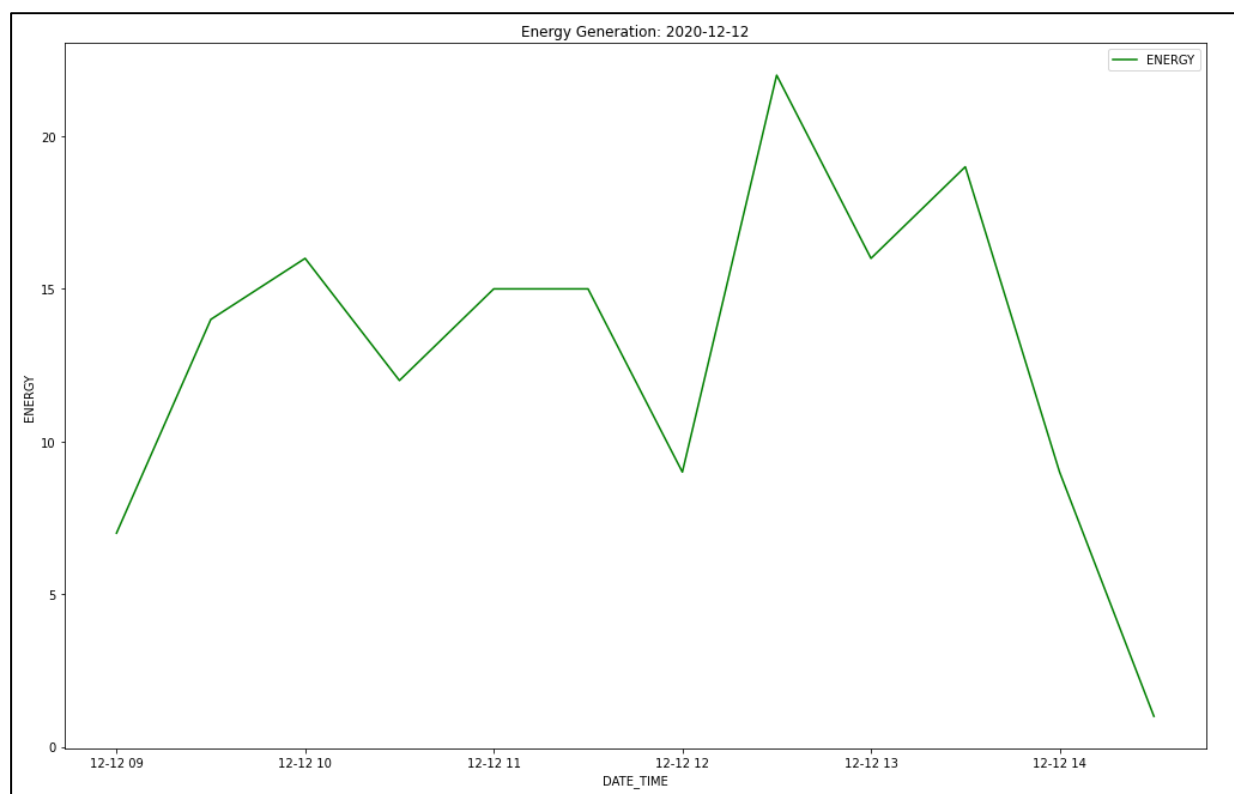


Рисунок 2.28 – Добовий графік електрогенерації за 12.12.2021
з присутньою аномалією

На рисунку 2.29 наведено фрагмент коду пошуку та фільтрації даних з наявними у датасеті аномальними значеннями електрогенерації.

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

	ENERGY	CLOUDY	YEAR	MONTH	DAY	TOTAL_DAY	TOTAL_TIME
count	15371.000000	15371.000000	15371.000000	15371.000000	15371.000000	15371.000000	15371.000000
mean	358.514280	38.616876	2020.597229	6.359313	15.746601	178.276625	753.554095
std	303.345812	35.117235	0.586857	3.001220	8.746692	91.702924	210.586309
min	11.000000	0.000000	2020.000000	1.000000	1.000000	1.000000	330.000000
25%	84.000000	0.000000	2020.000000	4.000000	8.000000	106.000000	570.000000
50%	261.000000	30.000000	2021.000000	6.000000	16.000000	177.000000	750.000000
75%	615.000000	70.000000	2021.000000	9.000000	23.000000	248.000000	900.000000
max	1131.000000	100.000000	2022.000000	12.000000	31.000000	366.000000	1230.000000

Рисунок 2.29 – Таблиці відфільтрованих даних

Щоб проаналізувати залежність зміни генерації від пори року, побудуємо графік, що відображає пік генерації електроенергії за час, що дорівнює часу дискретизації даних. Фрагмент коду (рис. 2.30).

```
# Data visualization
daily_dc = df_solar.groupby('MONTH')['ENERGY'].agg('max')
|
ax = daily_dc.plot.bar(figsize=(20,5), legend=True,color='red')
plt.title('Daily Energy')
plt.show()
```

Рисунок 2.30 – Побудова графіка місячної генерації

Як бачимо з рисунка 2.31, пік генерації припадає на весняні та літні місяці і обумовлено це тим, що сонце в цей період розташоване у найвищій точці по відношенню до горизонту.

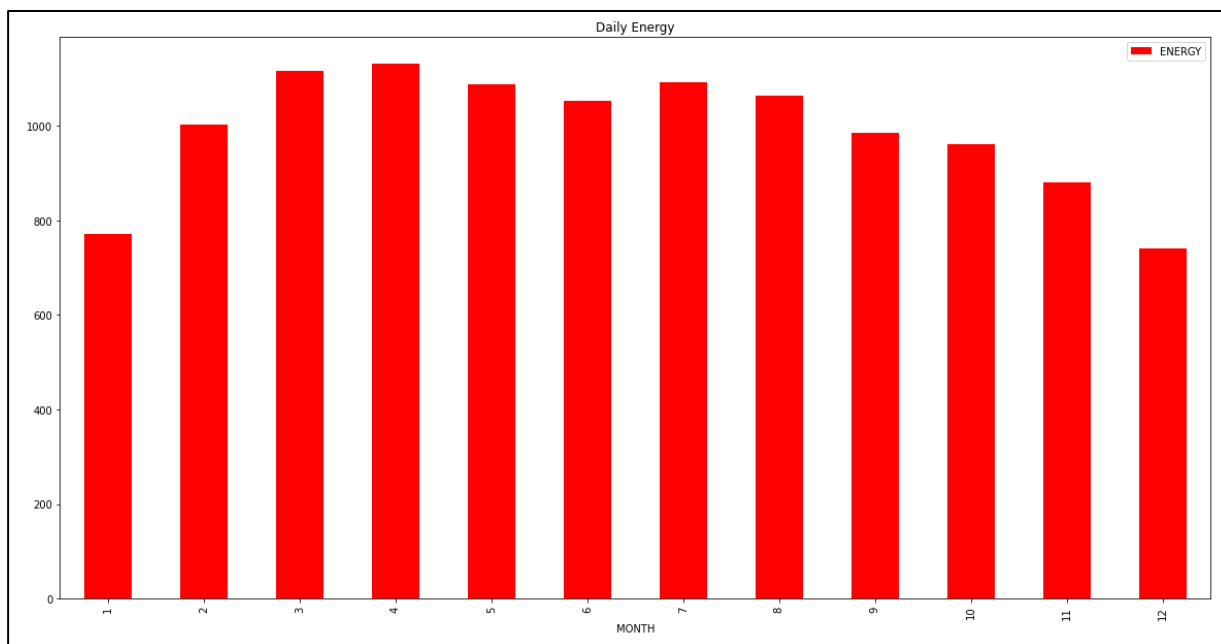


Рисунок 2.31 – Графік усередненої місячної генерації

Проаналізувавши наведені графіки та рисунок 2.32 приходимо до висновку, що кількість електроенергії, яка генерується сонячними панелями напряму залежить від [11]:

- погодних умов, а саме хмарності;
- тривалості сонячного дня та розташування zenіту відносно горизонту, тобто має сезонну залежність і напряму залежить від місяця року;
- час визначає точне положення сонця відносно горизонту.

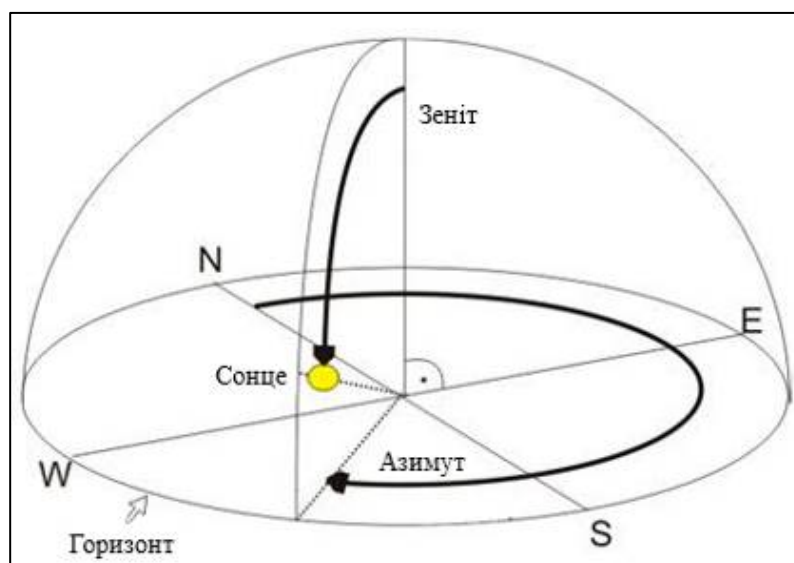


Рисунок 2.32 – Рух сонця відносно zenіту та азимуту

Відповідно стовпчик «ENERGY» і є той самий «таргет», який ми хочемо передбачити за даними інших показників, а саме «CLOUDY», «TOTAL_DAY», «TOTAL_TIME».

Дослідимо, як розподілені похмури і сонячні дні протягом року. Для цього побудуємо графік, хмарності в розрізі року. Фрагмент коду (рис. 2.33) та результат його роботи (рис. 2.34).

```
daily_cloudy = df_solar.groupby('MONTH')['CLOUDY'].agg('median')  
  
daily_cloudy.plot.bar(figsize=(20,10), legend=True,color='blue')  
plt.title('Cloudy')  
plt.show()
```

Рисунок 2.33 – Побудова графіка хмарності в розрізі року

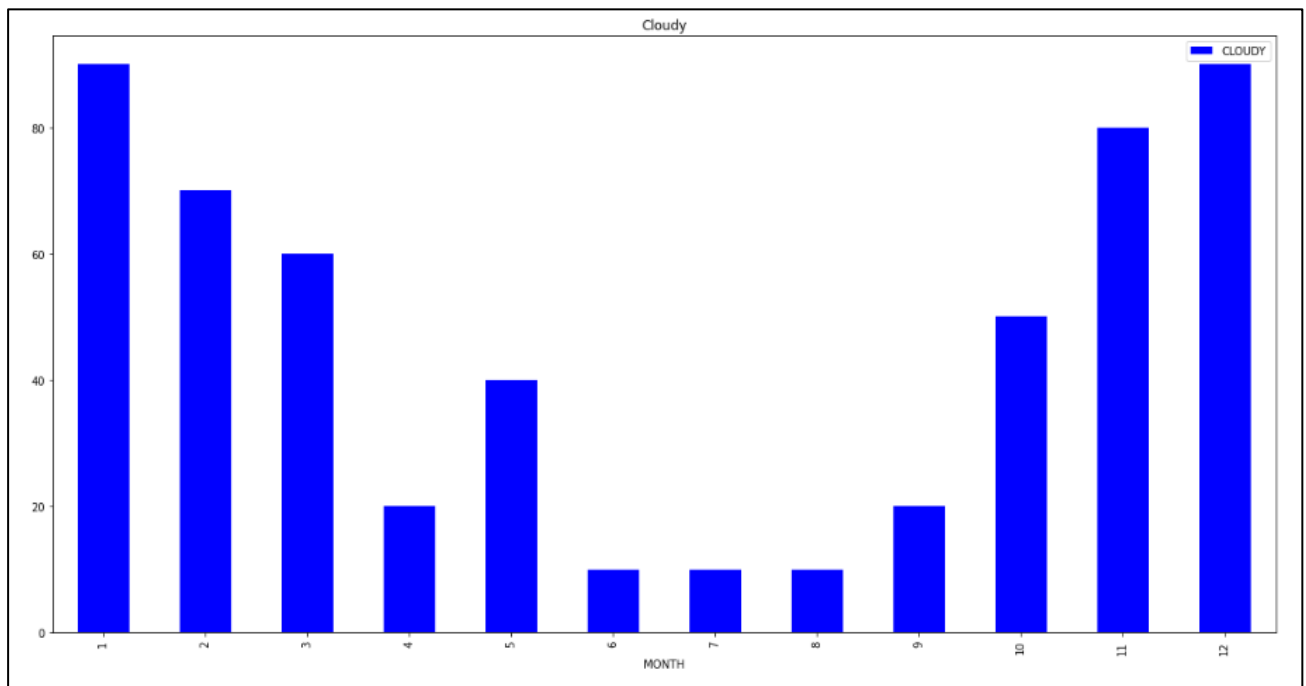


Рисунок 2.34 – Графік хмарності в розрізі року

2.4 Розвідувальний аналіз

Для здійснення розвідувального аналізу, вхідні дані перетворюють або відображають у зручній для сприйняття формі, що може бути правильним та коректним. Результати даного аналізу часто відображаються у формі схем, діаграм, графіків, таблиць та інших візуальних елементів. Ці методи аналізу часто включають:

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

Результати розвідувального аналізу зазвичай не використовуються безпосередньо для прийняття управлінських рішень. Вони слугують підґрунтям при розробці оптимальної стратегії, поглибленого вивчення або застосування математичних методів для аналізу, висування гіпотез і подальшого дослідження [32].

Використаємо «ProfileReport» за для побудови та відображення різноманітних графіків, гістограм по кожному показнику. За допомогою яких можна виконати розвідувальний аналіз (рис. 2.35).

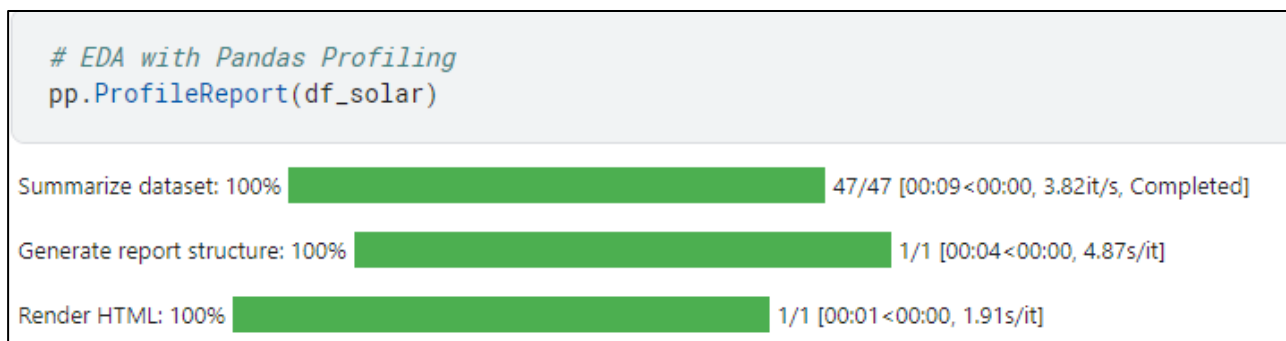


Рисунок 2.35 – Результат роботи «ProfileReport»

Результат виконання цієї команди, відображено у вигляді гістограми даних по кожному з показників, які будуть використані при аналізі і подальшому передбаченні процесу генерації електроенергії на фотоелектростанції (рис. 2.36).

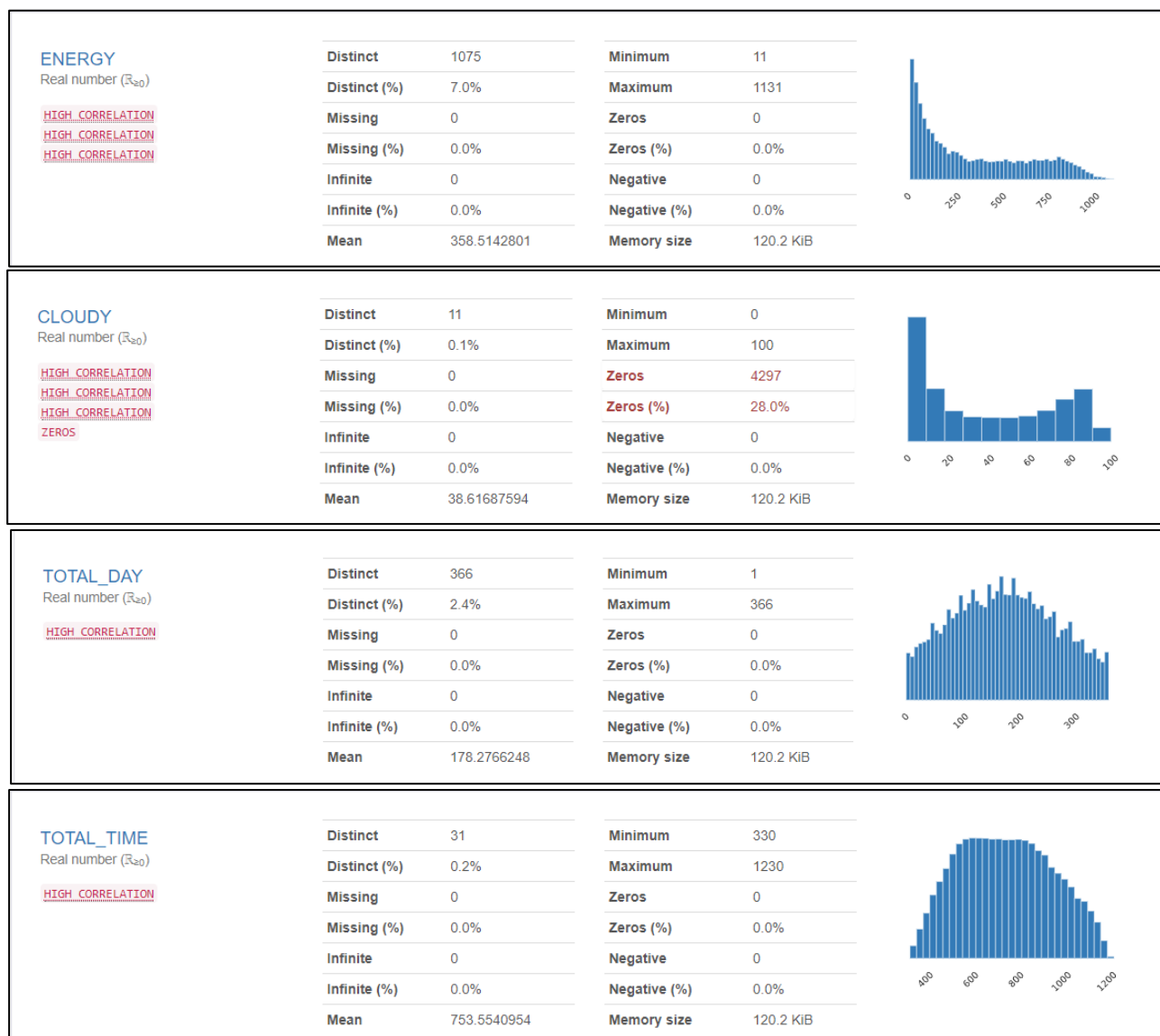


Рисунок 2.36 – Гістограми показників

Однією з особливостей «Pandas Profiling Report» є вкладка «Interactions» в якій візуально відображено кореляцію між показниками (рис. 2.37).

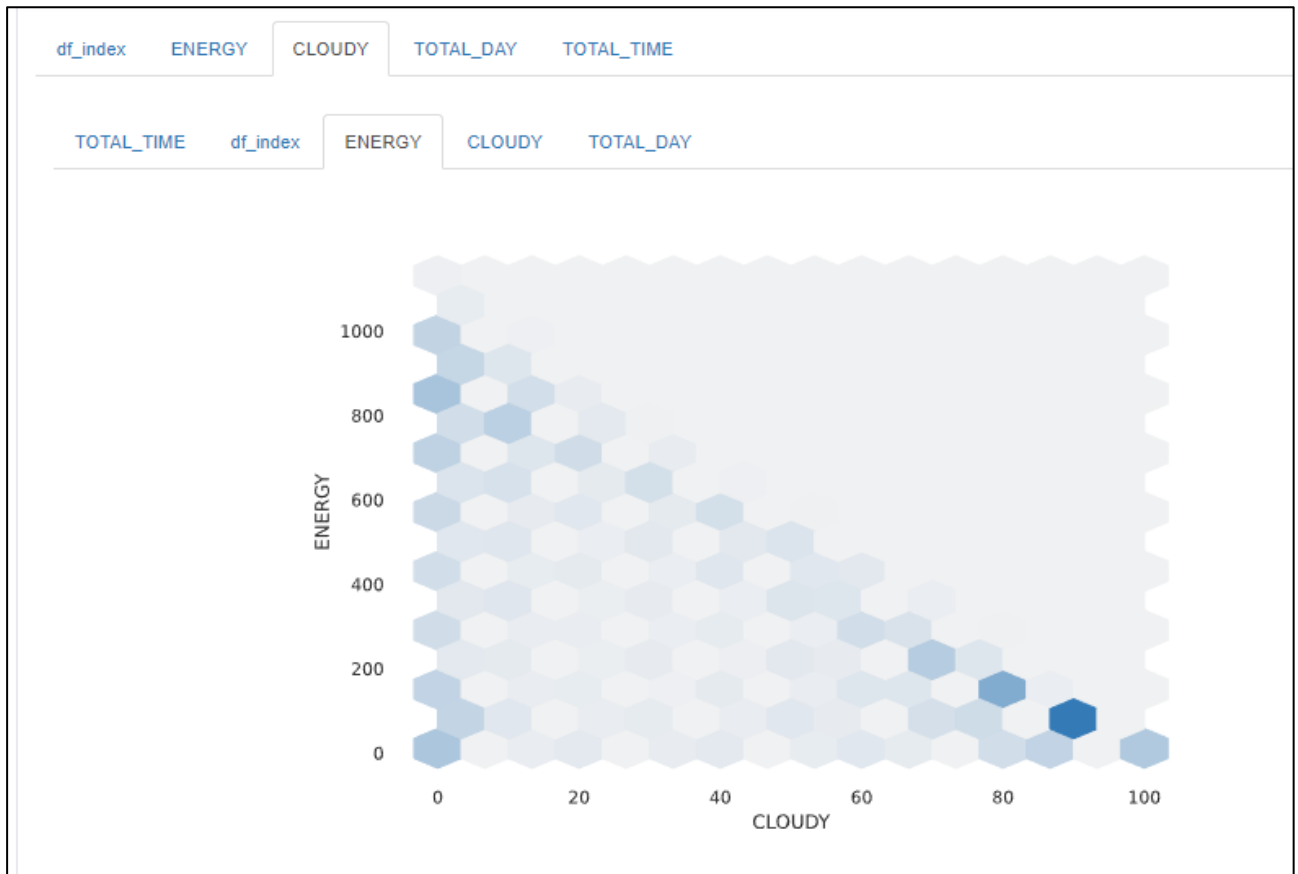


Рисунок 2.36 – Кореляція між «ENERGY» та «CLOUDY»

Тобто можна оцінити які аномалії, наскільки вони близькі до лінійної регресії, як показано на рисунку 2.37. В такий спосіб можна зробити попередні висновки співвідношень між даними показників [33].

Коефіцієнт кореляції Пірсона (r) вимірює ступінь лінійної залежності між двома змінними. Значення цього коефіцієнта може знаходитися в діапазоні від -1 до $+1$: -1 показує загальну негативну лінійну залежність, 0 вказує на відсутність лінійної зв'язку, а 1 вказує на загальну позитивну лінійну залежність [34]. Незалежно від змін розташування або масштабу двох змінних, значення r залишається однаковим. Наприклад, для лінійної функції, кут від осі x не впливає на r (рис. 2.38).

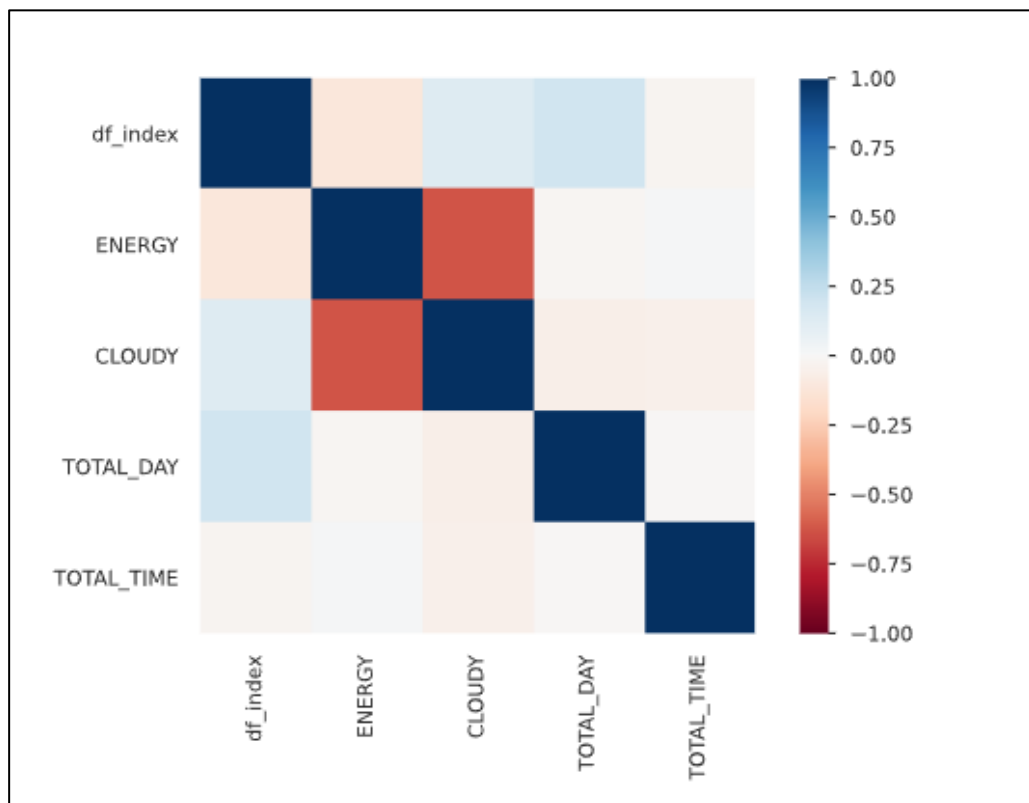


Рисунок 2.38 – Матриця кореляції Пірсона

Коефіцієнт рангової кореляції Спірмена (ρ) вимірює монотонну залежність між двома змінними і вважається більш відповідним для виявлення нелінійних монотонних зв'язків, ніж коефіцієнт кореляції Пірсона. Це значення також може знаходитися в діапазоні від -1 до +1: -1 вказує на загальну негативну монотонну залежність, 0 вказує на відсутність монотонної зв'язку, а 1 вказує на загальну позитивну монотонну залежність (рис. 2.39).

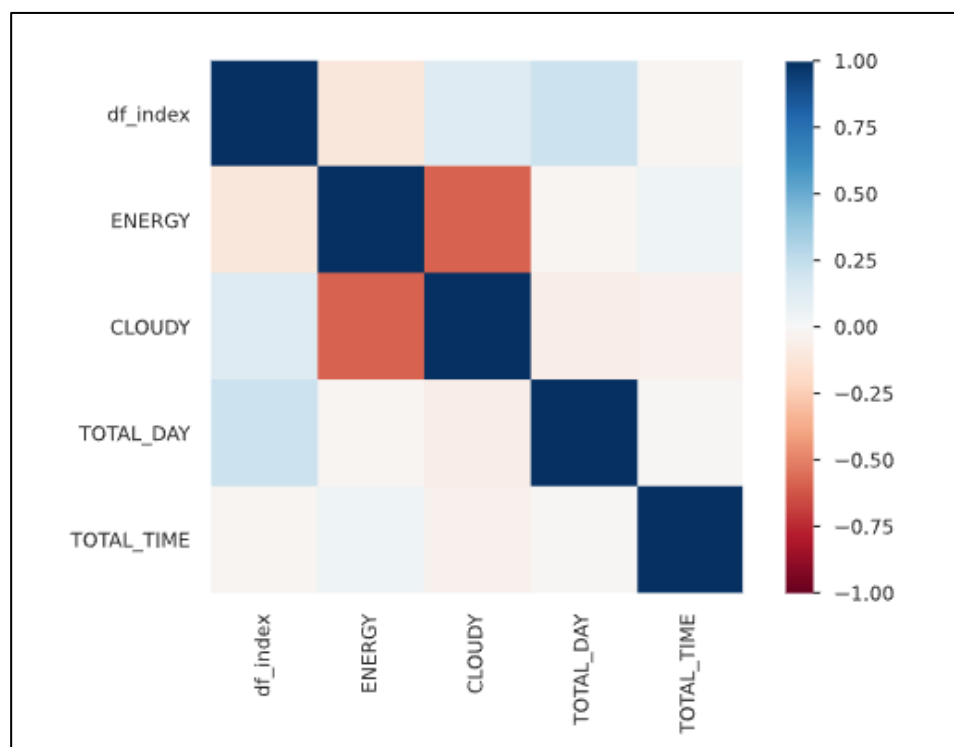


Рисунок 2.39 – Матриця кореляції Спірмена

Коефіцієнт кореляції рангу Кендалла (τ) вимірює порядковий зв'язок між двома змінними, схоже до коефіцієнта Спірмена. Це значення також може лежати в діапазоні від -1 до +1: -1 вказує на загальну негативну кореляцію, 0 вказує на відсутність кореляції, а 1 вказує на загальну позитивну кореляцію (рис. 2.40).

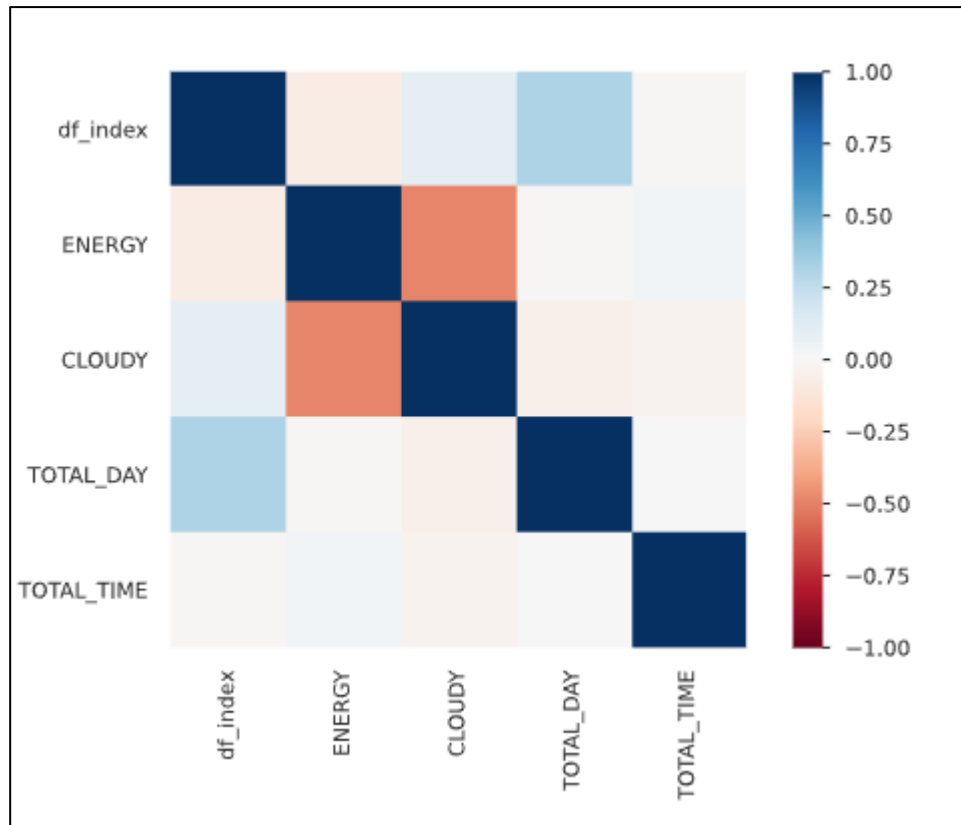


Рисунок 2.40 – Матриця кореляції Кендалла

Коефіцієнт кореляції Φ_{ik} (ϕ_k) є новим і практичним показником, який співвідносить категоріальні, порядкові та інтервальні змінні, зберігаючи послідовність. Цей коефіцієнт фіксує нелінійну залежність і, у разі двовимірного нормального розподілу вхідних даних, повертається до коефіцієнта кореляції Пірсона (рис. 2.41).

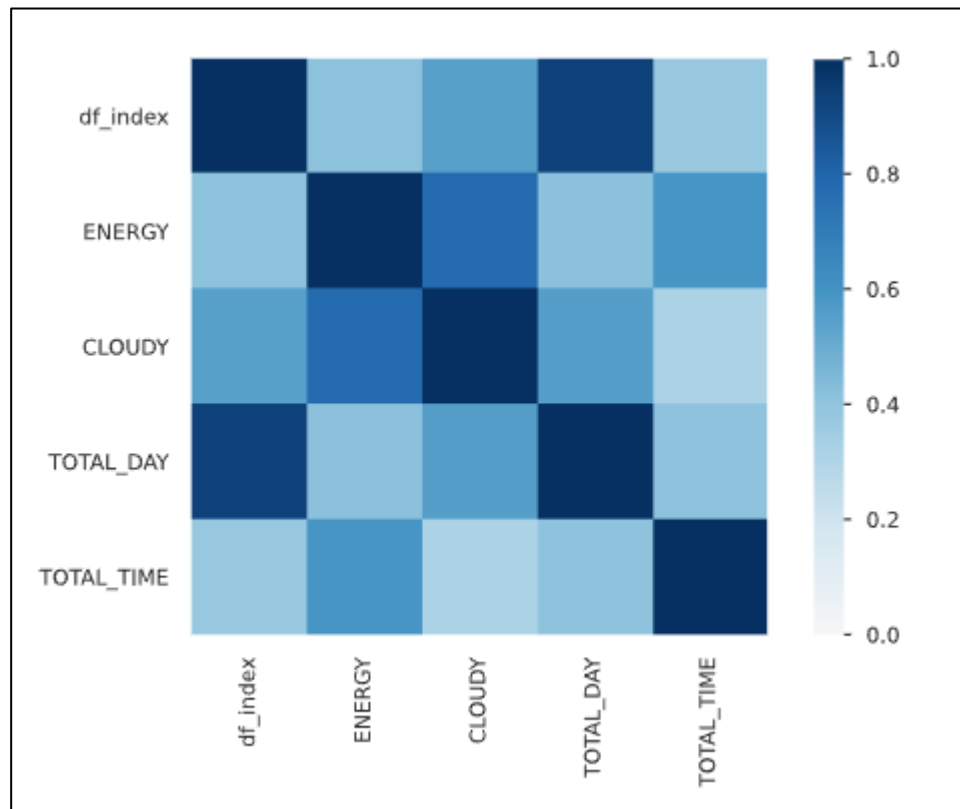


Рисунок 2.41 – Матриця кореляції Φ_{ik}

Як ми можемо бачити з матриці Пірсона, Спірмена, Кендалла кореляція між даних вимірюється за шкалою від -1 до 1. Виконуючи аналіз кореляцій між змінними, можемо помітити, що фактор “ENERGY” корелює на 60% із цільовою змінною “CLOUDY” [19-24].

Наступною є діаграма пропущених даних, але по причині початкової фільтрації даних всі стовпці мають значення у комірці. Що відображено на рисунку 2.42.

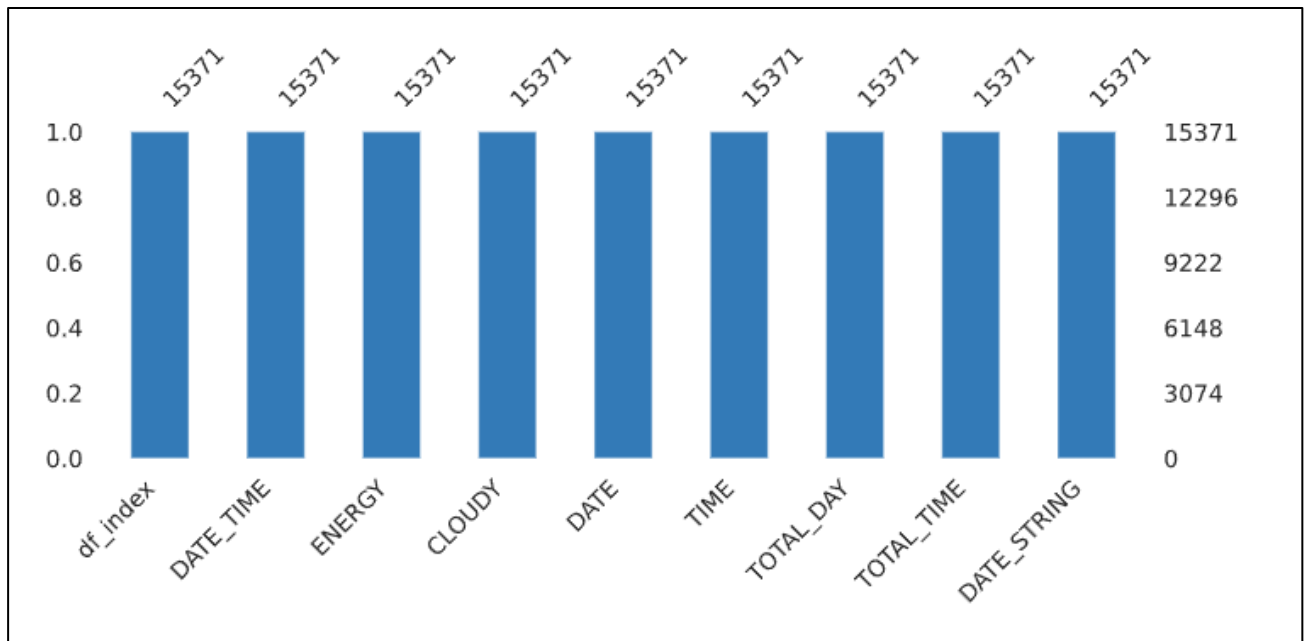


Рисунок 2.42 – Діаграма кількості пропущених даних

Сформуємо та розіб'ємо наші дані на тренувальні та тестові за допомогою команди «train_test_split», обсяг тестових даних приймаємо 40%, (рис. 2.43).

```
df = df_solar.copy()
X = df[['TOTAL_DAY', 'TOTAL_TIME', 'CLOUDY']]
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 2.43 – Розподіл даних

Статистика та кількість даних, що розподілилась між тренувальним та тестовим «датасетом» відображена на рисунках 2.44, 2.45.

X_train.describe()			
	TOTAL_DAY	TOTAL_TIME	CLOUDY
count	9222.000000	9222.000000	9222.000000
mean	178.361093	753.324658	38.233572
std	91.245616	210.907086	35.075213
min	1.000000	330.000000	0.000000
25%	106.000000	570.000000	0.000000
50%	177.000000	750.000000	30.000000
75%	247.000000	900.000000	70.000000
max	366.000000	1230.000000	100.000000

Рисунок 2.44 – Статистика тренувальних даних

X_test.describe()			
	TOTAL_DAY	TOTAL_TIME	CLOUDY
count	6149.000000	6149.000000	6149.000000
mean	178.149943	753.898195	39.191738
std	92.391826	210.120972	35.175189
min	1.000000	330.000000	0.000000
25%	106.000000	570.000000	0.000000
50%	177.000000	750.000000	30.000000
75%	249.000000	930.000000	70.000000
max	366.000000	1230.000000	100.000000

Рисунок 2.45 – Статистика тестових даних

У свою чергу, тренувальні дані ще раз необхідно розділити, виділити з них валідаційні. Ця частина «датасету» необхідна для вибору найкращої моделі, після чого слід застосувати її до тестових даних. [25, 26]. Тут валідаційні дані вибираються у розмірі 20% від загальної кількості тренувальних даних, (рис.2.46).

```
# Training data splitting to new training (part of the all training) and validation data
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 2.46 – Вибірка валідаційних даних

Згідно підготовлених даних можна приступати до подальшої побудови моделі, яка буде здатна передбачати об'єм генерації активної енергії, по доведеному прогнозу погоди (хмарності) за зазначений проміжок часу.

2.5 Висновки

Провівши попередній аналіз даних, виявлено аномалії та зашумленість у наявному датасеті. Після їх фільтрації була створена додаткова вибірка, в якій вже відсутні ці вказані проблеми. Використання цієї оновленої вибірки сприятиме значному підвищенню точності прогнозування.

Здійснено підготовку даних та виконано розвідувальний аналіз факторів. Для числових ознак, побудовано кореляційні матриці. Аналізуючи кореляцію між різними факторами, ми можемо бачити, що фактор хмарності корелює на 60% із цільовою змінною – електрогенерацією.

Підготовлено дані для подальшої побудови досліджуваних моделей.

3 СИСТЕМНИЙ АНАЛІЗ ТА ПЕРЕДБАЧЕННЯ ДОБОВОГО ГЕНЕРУВАННЯ ЕЛЕКТРОЕНЕРГІЇ ФОТОЕЛЕКТРОСТАНЦІЄЮ

3.1 Модель об'єкта

По причині того, що у попередніх розділах було визначено та підготовлено всі дані для подальшого передбачення добового генерування електроенергії фотоелектростанцією з урахуванням хмарності. Приступимо безпосередньо до виконання цієї задачі методом Decision Tree Regressor, Random Forest Regressor, LightGBM Regressor, що зображено на рисунку 3.1. По завершенню виберемо найкращу модель із трьох перерахованих.

```
# Creation the dataframe with the resulting score of all models
result = pd.DataFrame({'model' : ['Decision Tree Regressor', 'Random Forest Regressor', 'LightGBM Regressor'],
                      'train_score': 0, 'valid_score': 0})
result
```

	model	train_score	valid_score
0	Decision Tree Regressor	0	0
1	Random Forest Regressor	0	0
2	LightGBM Regressor	0	0

Рисунок 3.1 – Підготовка кадру даних з отриманою оцінкою усіх моделей

Виконаємо тренування та проаналізуємо результати тренувальних даних, які були передбачені за допомогою «Decision Tree Regressor», (рис. 3.2).

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

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

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

# Save to result dataframe
result.loc[result['model'] == 'Decision Tree Regressor', 'train_score'] = r2_score_dtr

print(f'Accuracy of Decision Tree Regressor model training is {r2_score_dtr}')

Accuracy of Decision Tree Regressor model training is 0.9997547534287423
```

Рисунок 3.2 – Точність передбачення навчальної моделі «Decision Tree Regressor»

В даному випадку ми бачимо, що на тренувальних даних 0,9998 а на валідаційних 0,9848, (рис. 3.3).

```

y_val_dtr = dtr.predict(X_valid)
r2_score_dtr_valid = r2_score(y_valid['ENERGY'], y_val_dtr)
result.loc[result['model'] == 'Decision Tree Regressor', 'valid_score'] = r2_score_dtr_valid
print(f'Accuracy of Decision Tree Regressor model prediction for valid dataset is {r2_score_dtr_valid}')

```

Accuracy of Decision Tree Regressor model prediction for valid dataset is 0.9848411838279288

Рисунок 3.3 – Точність передбачення дійсної моделі
«Decision Tree Regressor»

Виконаємо тренування та проаналізуємо результати тренувальних даних, які були передбачені за допомогою «Random Forest Regressor», (рис. 3.4).

```

# Random Forest Regressor
rfr = RandomForestRegressor()

# Training model
rfr.fit(X_train,y_train)

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

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

# Save to result dataframe
result.loc[result['model'] == 'Random Forest Regressor', 'train_score'] = r2_Score_rfr

print(f'Accuracy of Random Forest Regressor model training is {r2_Score_rfr}')

```

Accuracy of Random Forest Regressor model training is 0.9982492460185914

Рисунок 3.4 – Точність передбачення навчальної моделі
«Random Forest Regressor»

Як можна бачити, на тренувальних даних отримали результат 0,9982, а на валідаційних 0,9908, (рис. 3.5).

```

y_val_rfr = rfr.predict(X_valid)
r2_score_rfr_valid = r2_score(y_valid['ENERGY'], y_val_rfr)
result.loc[result['model'] == 'Random Forest Regressor', 'valid_score'] = r2_score_rfr_valid
print(f'Accuracy of Random Forest Regressor for valid dataset is {r2_score_rfr_valid}')

```

Accuracy of Random Forest Regressor for valid dataset is 0.9907789308094831

Рисунок 3.5 – Точність передбачення дійсної моделі
«Random Forest Regressor»

Виконаємо тренування та проаналізуємо результати тренувальних даних які були передбачені за допомогою «LightGBM Regressor», (рис. 3.6).

```

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 data
y_pred_lgb = model_lgb.predict(X_train)

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

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

print(f'Accuracy of LightGBM Regressor model training is {r2_Score_lgb}')

```

[LightGBM] [Warning] Auto-choosing row-wise multi-threading, the overhead of testing was 0.001686 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 295

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

[LightGBM] [Info] Start training from score 364.811577

Accuracy of LightGBM Regressor model training is 0.996874292114864

Рисунок 3.6 – Точність передбачення навчальної моделі
«LightGBM Regressor»

Як можна бачити, на тренувальних даних отримали результат 0,9969, а на валідаційних 0,9925, (рис. 3.7).

```

y_val_lgb = model_lgb.predict(X_valid)
r2_score_lgb_valid = r2_score(y_valid['ENERGY'], y_val_lgb)
result.loc[result['model'] == 'LightGBM Regressor', 'valid_score'] = r2_score_lgb_valid
print(f'Accuracy of LightGBM Regressor for valid dataset is {r2_score_lgb_valid}')

```

Accuracy of LightGBM Regressor for valid dataset is 0.9925139383127978

Рисунок 3.7 – Точність передбачення дійсної моделі
«LightGBM Regressor»

Візуалізуємо передбачення навчальних валідаційних і тестових даних.

Фрагменти коду, для побудови графіків наведено на рисунках 3.8-3.10.

```

# Prediction of target for training data for all models
shift = 2000
count = 100
index = X_train.index[shift:shift+count]
X_train = X_train.loc[index]
y_train = y_train.loc[index]

y_pred_dtr = dtr.predict(X_train)
y_pred_rfr = rfr.predict(X_train)
y_pred_lgb = model_lgb.predict(X_train)

x = X_train["TOTAL_TIME"][index];

# Building plot for prediction for the training data
plt.figure(figsize=(20,15))
plt.scatter(x, y_train, label = "Target training data", color = 'g', marker='^')
plt.scatter(x, y_pred_dtr, label = "Decision Tree Regressor prediction", color = 'b', marker='x')
plt.scatter(x, y_pred_rfr, label = "Random Forest prediction", color = 'y', marker='s')
plt.scatter(x, y_pred_lgb, label = "LightGBM Regressor prediction", color = 'brown', marker='o')
plt.plot(x, np.full(len(X_train), 1200), label = "Maximum allowable value", color = 'r')
plt.title('Prediction for the training data')
plt.legend(loc='upper right')
plt.grid(True)

```

Рисунок 3.8 – Код для побудови графіку передбачення навчальних даних

```

# Prediction of target for valid data for all models
shift = 1000
count = 100
index = X_valid.index[shift:shift+count]
X_valid = X_valid.loc[index]
y_valid = y_valid.loc[index]

y_val_dtr = dtr.predict(X_valid)
y_val_rfr = rfr.predict(X_valid)
y_val_lgb = model_lgb.predict(X_valid)

x = X_valid["TOTAL_TIME"][index];

# Building plot for prediction for the valid data
plt.figure(figsize=(20,15))
plt.scatter(x, y_valid, label = "Target valid data", color = 'g', marker='^')
plt.scatter(x, y_val_dtr, label = "Decision Tree Regressor prediction", color = 'b', marker='x')
plt.scatter(x, y_val_rfr, label = "Random Forest prediction", color = 'y', marker='s')
plt.scatter(x, y_val_lgb, label = "LightGBM Regressor prediction", color = 'brown', marker='o')
plt.plot(x, np.full(len(X_valid), 1200), label = "Maximum allowable value", color = 'r')
plt.title('Prediction for the valid data')
plt.legend(loc='upper right')
plt.grid(True)

```

Рисунок 3.9 – Код для побудови передбачення валідаційних даних

```

# Prediction of target for test data for all models
shift = 2000
count = 100
index = X_test.index[shift:shift+count]
X_test = X_test.loc[index]
y_test = y_test.loc[index]

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

x = X_test["TOTAL_TIME"][index];

# Building plot for prediction for the test data
plt.figure(figsize=(20,15))
plt.scatter(x, y_test, label = "Target test data", color = 'g', marker='^')
plt.scatter(x, y_test_dtr, label = "Decision Tree Regressor prediction", color = 'b', marker='x')
plt.scatter(x, y_test_rfr, label = "Random Forest prediction", color = 'y', marker='s')
plt.scatter(x, y_test_lgb, label = "LightGBM Regressor prediction", color = 'brown', marker='o')
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 3.10 – Код для побудови передбачення тестових даних

Графіки прогнозування навчальних, валідаційних і тестових даних зображені на рисунках 3.11-3.13 відповідно. Червоною лінією зображена встановлена (максимально можлива) електрогенерація фотоелектростанцією.

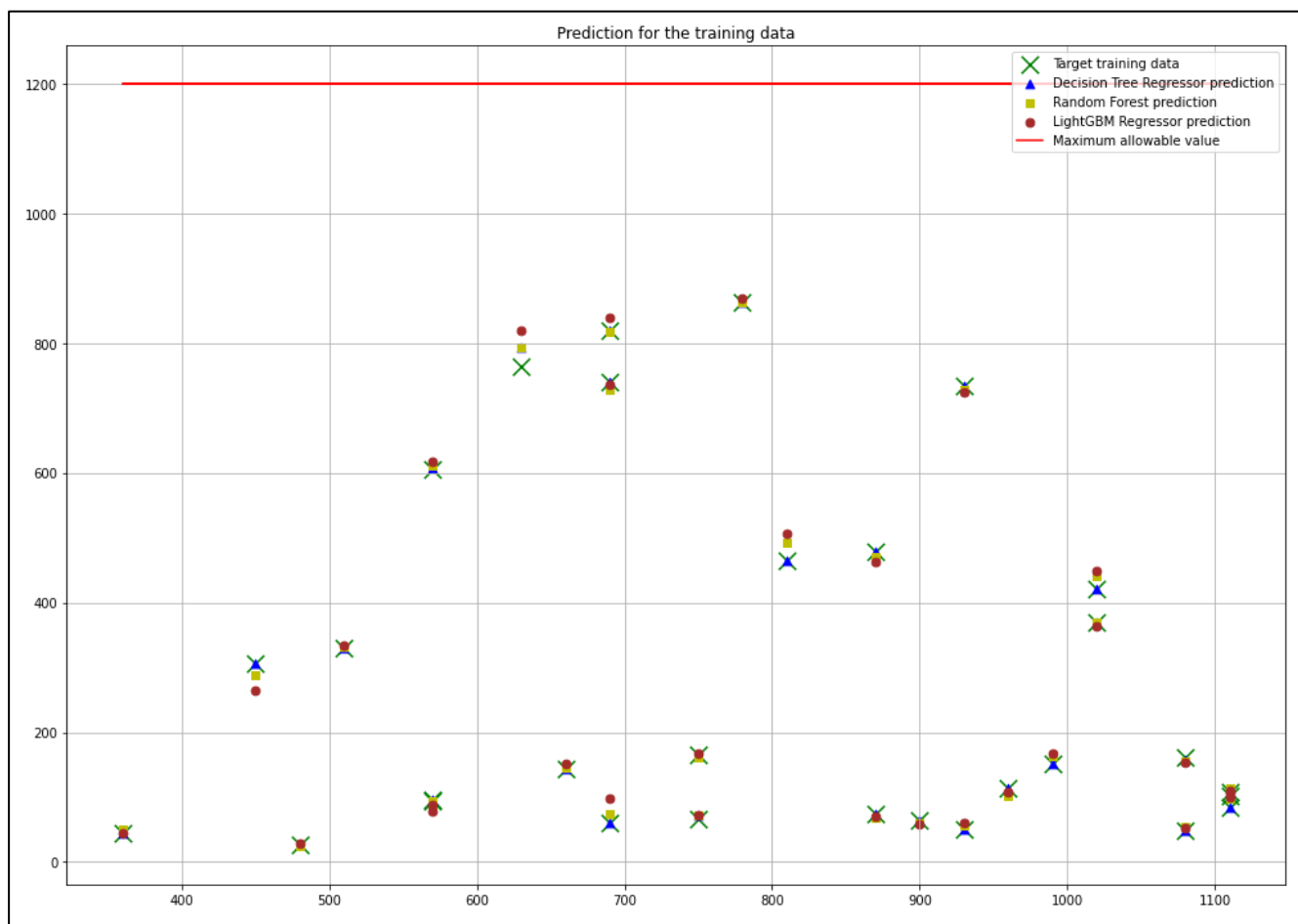


Рисунок 3.11 – Передбачення навчальних даних

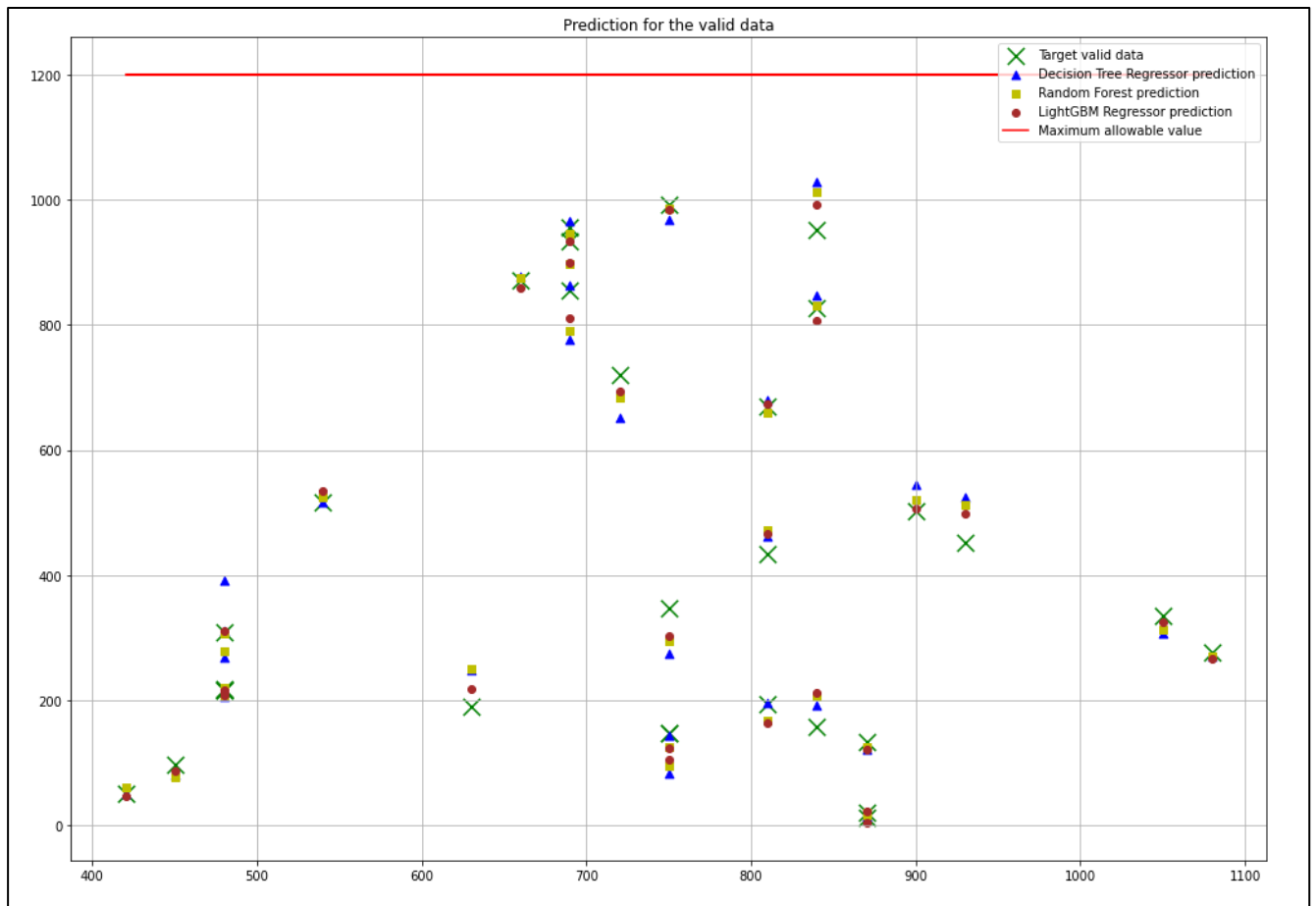


Рисунок 3.12 – Передбачення валідаційних даних

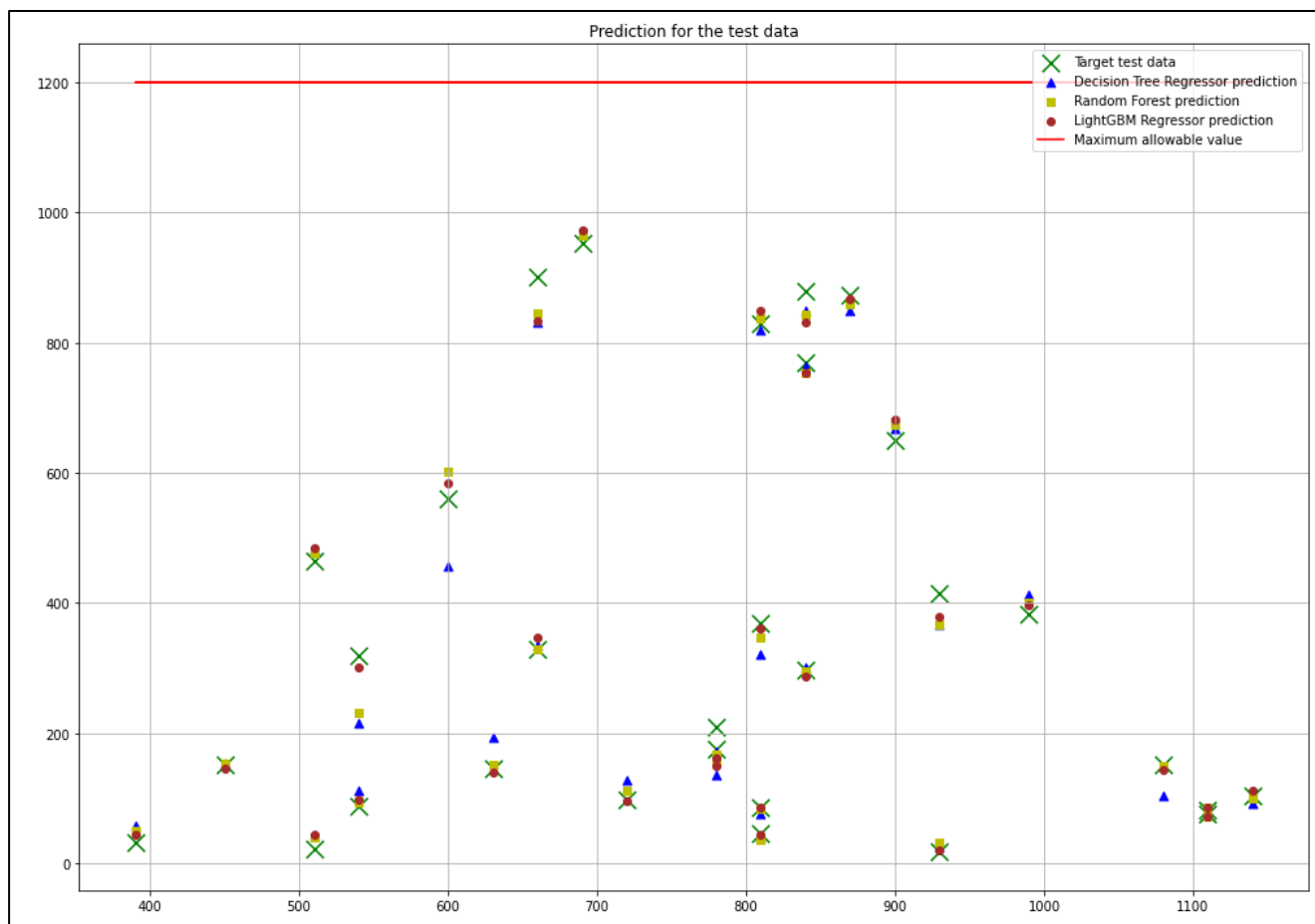


Рисунок 3.13 – Передбачення тестових даних

Тепер, коли всі розрахунки виконані, можна приступати до наступного етапу – вибору найкращої моделі прогнозування. На рисунку 3.14 відображено таблицю з результатами передбачення, що надають досліджувані моделі.

```
# Display results of modeling
result.sort_values(by=['valid_score', 'train_score'], ascending=False)
```

	model	train_score	valid_score
2	LightGBM Regressor	0.996874	0.992514
1	Random Forest Regressor	0.998249	0.990779
0	Decision Tree Regressor	0.999755	0.984841

Рисунок 3.14 – Таблиця результатів передбачення

Важливо, щоб різниця між отриманими показниками була не занадто великою. Першою вибирається найкраща позиція, для якої різниця між тренувальними і валідаційними даними не перевищує 10%, а далі між ними вибирається та у якої найбільший «valid_score», (рис. 3.15).

```
# Select models
result_best = result[(result['train_score'] - result['valid_score']).abs() < 0.1]
result_best.sort_values(by=['valid_score', 'train_score'], ascending=False)
```

	model	train_score	valid_score
2	LightGBM Regressor	0.996874	0.992514
1	Random Forest Regressor	0.998249	0.990779
0	Decision Tree Regressor	0.999755	0.984841

Рисунок 3.15 – Вибір моделі з мінімальним відхиленням

Наступним кроком є відображення найкращого з результатів передбачення. Фрагмент коду та результат його роботи наведені на рисунку 3.16.

```
# Find a name of the best model (with maximal valid score)
best_model_name = result_best.loc[result_best['valid_score'].idxmax(), 'model']
print(f'The best model is "{best_model_name}")
```

The best model is "LightGBM Regressor"

Рисунок 3.16 – Найкраща модель передбачення

Отже, виходячи з результату роботи алгоритму, найкраща модель передбачення була побудована на основі метода «LightGBM Regressor».

3.2 Висновки

Побудовано ряд моделей прогнозування генерації електроенергії з використанням різних методів. Було проведено навчання моделей, перевірку їх точності на різних наборах даних та побудову візуальних діаграм. Найкращою моделлю, згідно метрики $r2_score$ 0,99 на валідаційному наборі, була обрана модель на основі методу "LightGBM". Навчена модель здатна з великою точністю передбачати добову генерацію електроенергії, що буде мати вагомий економічний ефект.

4 ЕКОНОМІЧНА ЧАСТИНА

4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки

Ефективне управління генерацією електроенергії з використанням відновлюваних джерел стає насущною проблемою у багатьох країнах, включаючи Україну. Річне зростання генерації відновлюваних джерел енергії, зокрема сонячної та вітрової, призводять до нестабільності в електроенергетичній системі через залежність від погодних умов.

Передбачення генерації електроенергії від цих джерел стає ключовим для збалансованого розподілу виробництва та споживання, а також для формування відповідних цін на електроенергію. Це важливо для забезпечення стабільності енергетичної системи та ефективного використання відновлюваних джерел.

Особлива увага приділяється аналізу та передбаченню генерації сонячної енергії. Комерційний аудит оцінює економічні та фінансові аспекти науково-технічних розробок, визначає їх вплив на ринок та стратегічні переваги для бізнесу. Це включає аналіз генерації сонячної енергії, як ключового параметра для визначення рентабельності фотоелектростанцій.

Технологічний аспект аудиту включає оцінку існуючих та майбутніх технологічних рішень у сфері сонячної енергії, підвищення ефективності сонячних панелей, систем зберігання енергії та моніторингу. Передбачення генерації стає важливим параметром для оцінки технічної ефективності новацій та їхнього впровадження на ринок.

Аудит у сфері відновлювальних джерел енергії є критично важливим у сучасному енергетичному ландшафті, враховуючи стратегічне значення генерації електроенергії від сонячних джерел для сталого енергетичного майбутнього.

Критерії оцінювання наукового-технічного рівня та комерційного потенціалу розробки наведено в таблиці 4.1 [35].

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
1	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 аналогівна малому ринку	Мало аналогівна малому ринку	Кілька аналогівна великому ринку	Один аналог навеликому ринку	Продукт не маєаналогів на великому ринку
3	Ціна продуктузначно вища заціни аналогів	Ціна продуктудещо вища заціни аналогів	Ціна продуктуприблизно дорівнює цінам аналогів	Ціна продуктудещо нижча заціни аналогів	Ціна продуктузначно нижча заціни аналогів
4	Технічні таспоживчі властивості продукту значно гірші,ніж в аналогів	Технічні таспоживчі властивості продукту трохи гірші,ніж в аналогів	Технічні таспоживчі властивості продукту порівніні аналогів	Технічні таспоживчі властивості продукту трохи кращі,ніж в аналогів	Технічні таспоживчі властивості продуктузначно кращі,ніж в аналогів
5	Експлуатаційні витрати значновищі, ніж в аналогів	Експлуатаційні витрати дещовищі, ніж в аналогів	Експлуатаційні витрати на рівніексплуатаційнихвитрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значнонижчі, ніж в аналогів
<i>Ринкові перспективи</i>					
6	Ринок малий іне має позитивної динаміки	Ринок малий,але має позитивну динаміку	Середній ринок зпозитивною динамікою	Великий стабільний ринок	Великий ринокз позитивноюдинамікою
7	Активна конкуренція великихкомпаній наринку	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкурентівне має

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, що потребує значних коштів та часу	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 [35].

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. [35].

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[35]:

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

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, $T_p=21 \dots 23$ дні.

Проведені розрахунки зведено 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 проєкту	35200	1600	50	80000
Програміст	30800	1400	50	70000
Всього				150000

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 [35];

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. [35];

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

t_{3M} – тривалість зміни, год.

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

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

Проведені розрахунки зведено до 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.

$$Z_{\text{дод}} = (150000 + 2094,83) \cdot \frac{10}{100\%} = 15209,48 \text{ (грн.)}$$

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:

$$Z_n = (Z_o + Z_p + Z_{\text{дод}}) \cdot \frac{N_{\text{зп}}}{100\%}, \quad (4.5)$$

де $H_{зп}$ – норма нарахування на заробітну плату.

$$З_н = (150000 + 2094,83 + 15209,48) \cdot \frac{22}{100\%} = 36806,95 \text{ (грн).}$$

4.2.3 Сировина та матеріали

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

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

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;

$Ц_{vj}$ – 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	629,0	1,0	-	-	723,35
Офісний папір А4 500	215,0	3,0	-	-	741,75
Картридж для принтера	960,0	1,0	-	-	1104
Набір канцелярський офісний	330,0	2,0	-	-	759
Органайзер офісний	510,0	3,0	-	-	1759,50
Папір для записів А5	100,0	3,0	-	-	345
Всього					5432,60

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 та наукових (експериментальних) робіт

Витрати на програмне забезпечення на 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.

$$A_{\text{обл}} = \frac{28500 \cdot 3}{2 \cdot 12} = 3562,5 \text{ (грн)}.$$

Проведені розрахунки зведено в таблицю 4.7.

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 розробки та моделювання	28500	2	3	3562,50
Робоче місце інженера розробника	10600	5	3	530
Офісне приміщення	1200000	20	3	15000
ОС Windows 10 PRO	8000	2	3	1000
Прикладний пакет Microsoft Office 2021	11850	2	3	1481,25
Всього				21573,75

4.2.8 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію (B_e) розраховуємо за формулою:

$$B_e = \sum_{i=1}^n \frac{W_{yi} \cdot t_i \cdot \text{Ц}_e \cdot K_{\text{впн}i}}{\eta_i}, \quad (4.8)$$

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;

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

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

$$V_e = 0,35 \cdot 360,0 \cdot 7,50 \cdot 0,96 / 0,97 = 935,26 \text{ (грн).}$$

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	0,35	360,0	935,26
Робоче місце інженера	0,12	310,0	276,12
Всього			1211,38

4.2.9 Службові відрядження

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

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

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

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

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:

$$I_B = (Z_o + Z_p) \cdot \frac{H_{iB}}{100\%}, \quad (4.10)$$

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

$$I_B = (150000 + 2094,83) \cdot \frac{50}{100\%} = 76047,42 \text{ (грн)}.$$

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:

$$V_{H3B} = (Z_o + Z_p) \cdot \frac{H_{H3B}}{100\%}, \quad (4.11)$$

де H_{H3B} – норма нарахування за статтею «Накладні (загальновиробничі) витрати», прийmemo $H_{H3B} = 100\%$.

$$V_{\text{нзв}} = (150000 + 2094,83) \cdot 100 / 100\% = 152094,83 \text{ (грн)}.$$

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

$$V_{\text{заг}} = Z_0 + Z_P + Z_{\text{дод}} + Z_{\text{н}} + M + K_{\text{в}} + V_{\text{спец}} + V_{\text{прг}} + A_{\text{обл}} + V_{\text{е}} + V_{\text{св}} + V_{\text{сп}} + I_{\text{в}} + V_{\text{нзв}}, \quad (4.12)$$

$$V_{\text{заг}} = 150000 + 2094,83 + 15209,48 + 36806,95 + 5432,6 + 0 + 0 + 0 + 21573,75 + 1211,38 + 38023,7 + 0 + 76047,42 + 152094,83 = 498494,94 \text{ (грн)}.$$

Загальні витрати ЗВ 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 $\eta=0,90$.

$$ЗВ = \frac{498494,94}{0,90} = 553883,27 \text{ (грн)}.$$

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 сьогоднішні 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 [35]:

$$\Delta\Pi_i = (\pm\Delta C_0 \cdot N + C_0 \cdot \Delta N_i)_i \cdot \lambda \cdot \rho \cdot \left(1 - \frac{\vartheta}{100}\right), \quad (4.15)$$

де $\pm\Delta C_0$ – 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); $\pm\Delta C_0$ 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 500 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 25000 осіб;

C_0 – основний якісний показник, який визначає ціну реалізації нової науково-технічної розробки в аналізованому році, $C_0 = C_6 \pm \Delta C_0$;

C_6 – основний якісний показник, який визначає ціну реалізації існуючої (базової) науково-технічної розробки у році до впровадження результатів. Прийmemo 2500,00 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 $\rho = 0,3$;

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

Збільшення чистого прибутку 1-го року:

$$\Delta\Pi_1 = (250 \cdot 25000 + 2750 \cdot 3500) \cdot 0,83 \cdot 0,3 \cdot \left(1 - \frac{18\%}{100\%}\right) = 3241357,50 \text{ (грн)}.$$

Збільшення чистого прибутку 2-го року:

$$\Delta\Pi_2 = (250 \cdot 25000 + 2750 \cdot 8000) \cdot 0,83 \cdot 0,3 \cdot \left(1 - \frac{18\%}{100\%}\right) = 5768085,00 \text{ (грн)}.$$

Збільшення чистого прибутку 3-го року:

$$\Delta\Pi_3 = (250 \cdot 25000 + 2750 \cdot 14000) \cdot 0,83 \cdot 0,3 \cdot \left(1 - \frac{18\%}{100\%}\right) = 9137055,00 \text{ (грн)}.$$

Збільшення чистого прибутку 4-го року:

$$\Delta\Pi_4 = (250 \cdot 25000 + 2750 \cdot 16000) \cdot 0,83 \cdot 0,3 \cdot \left(1 - \frac{18\%}{100\%}\right) = 10260045,00 \text{ (грн)}.$$

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.

$$\text{ПП} = \frac{3241357,50}{(1+0,15)^1} + \frac{5768085,00}{(1+0,15)^2} + \frac{9137055,00}{(1+0,15)^3} + \frac{10260045,00}{(1+0,15)^4} = 2818571,74 + 4361500,95 + 6007761,98 + 5866214,03 = 19054048,69 \text{ (грн).}$$

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

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

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 $k_{\text{інв}} = 2,2$;
 $ЗВ$ – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, складають 553883,27 грн.

$$PV = k_{\text{інв}} \cdot ЗВ = 2,2 \cdot 553883,27 = 1218543,19 \text{ (грн).}$$

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:

$$E_{\text{абс}} = \text{ПП} - PV, \quad (4.18)$$

де ПП – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 19054048,69 грн;
PV – теперішня вартість початкових інвестицій, 1218543,19 грн.

$$E_{\text{абс}} = 19054048,69 - 1218543,19 = 17835505,50 \text{ (грн)}.$$

Внутрішня економічна дохідність інвестицій $E_{\text{в}}$, які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки, розраховується за формулою:

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

де $E_{\text{абс}}$ – абсолютний економічний ефект вкладених інвестицій, 17835505,50 грн;
PV – теперішня вартість початкових інвестицій, 1218543,19 грн;
 $T_{\text{ж}}$ – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримання позитивних результатів від її впровадження, 4 роки

$$E_{\text{в}} = \sqrt[4]{1 + \frac{E_{\text{абс}}}{PV}} - 1 = (1 + 17835505,50 / 1218543,19)^{1/4} - 1 = (15,64)^{1/4} - 1 = 1,99 - 1 = 0,99.$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій $\tau_{\text{мін}}$ 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 $f=0,05\dots0,5$, але може бути і значно вищою. Прийmemo 0,35.

$$\tau_{\text{мін}} = d + f = 0,16 + 0,35 = 0,51$$

$E_{\text{в}} > \tau_{\text{мін}}$, це свідчить про те, що внутрішня економічна дохідність інвестицій $E_{\text{в}}$, 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:

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

де $E_{\text{в}}$ – внутрішня економічна дохідність вкладених інвестицій.

$$T_{\text{ок}} = \frac{1}{0,99} = 1,01 \text{ р.}$$

Як бачимо термін окупності становить $T_{\text{ок}} < 3$ -х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок [35].

4.4 Висновки

Проведені дослідження підтверджують великий комерційний потенціал розробки, яка фокусується на аналізі та передбаченні добового виробництва електроенергії. Оцінка цього потенціалу становить 38 балів, що свідчить про значні можливості для успішної комерціалізації цього науково-технічного рішення.

Особливо обнадійливим є те, що відшкодування витрат на цю розробку відбудеться за 1,01 року. Це свідчить про високу комерційну привабливість та швидкий повернення інвестицій. Такий короткий термін відшкодування може стимулювати потенційного інвестора розглядати можливість фінансування впровадження цієї розробки та її введення на ринок.

Отже, з урахуванням отриманих результатів можна зробити висновок про доцільність науково-технічної роботи у рамках цієї теми. Успішна комерціалізація може призвести до значних успіхів на ринку та забезпечити інвестора високим рівнем чистого прибутку у найкоротший термін.

ВИСНОВКИ

Магістерська кваліфікаційна робота на тему: “Аналіз та передбачення добового генерування електроенергії фотоелектростанцією. Частина 2. Аналіз впливу хмарності”, присвячена застосуванню методів системного аналізу та машинного навчання для передбачення об’ємів генерації електроенергії з урахуванням впливу хмарності та мінімальною похибкою, що в свою чергу допоможе уникнути штрафів перед енергоринком та мінімізувати втрати, забезпечуючи стабільну поставку електроенергії.

Проведений аналіз генерації електроенергії на сонячних фотоелектричних станціях з урахуванням впливу хмарності виявив ряд важливих закономірностей. Виявлено, що інтенсивність сонячного випромінювання, яка є ключовим фактором у генерації електроенергії, піддається значним змінам через коливання хмарності.

Проаналізовано предмет досліджень, ретельно оглянуто наявні проблеми та їх актуальність, оцінено можливі шляхи їх вирішення. Детально проаналізовано основні методи передбачення електрогенерації з урахуванням впливу хмарності, які застосовуються у цій області, зокрема, оглянуті їх переваги та недоліки. Також розглянуто оптимальні інформаційні технології, спрямовані на впровадження методів машинного навчання для розв’язання цих проблем. Визначено, які конкретні технології та інструменти можуть бути найбільш ефективними в контексті розв’язання завдань прогнозування в даній сфері.

Проведено аналіз вхідних даних, включаючи побудову та аналіз добових графіків генерації електроенергії та хмарності. Також був проведений розвідувальний аналіз, в результаті якого було побудовано кореляційні матриці за методами Пірсона, Спірмана, Кендала, Фіка, за допомогою яких було визначено 60% кореляцію між цільовою змінною та фактором впливу.

Побудовано ряд моделей прогнозування генерації електроенергії з використанням різних методів. Було проведено навчання моделей, перевірку їх точності на різних наборах даних та побудову візуальних діаграм. Найкращою

моделлю, згідно метрики $r2_score$ 0,99 на валідаційному наборі, була обрана модель на основі методу "LightGBM". Навчена модель здатна з великою точністю передбачати добову генерацію електроенергії з урахуванням впливу хмарності, що буде мати вагомий економічний ефект.

Економічна частина даної роботи містить розрахунок витрат на розробку, сума яких складає 553883 гривень. Було спрогнозовано орієнтовану величину витрат по кожній з статей витрат. Розраховано чистий прибуток, який може отримати інвестор від реалізації нового технічного рішення, розраховано період окупності витрат та економічний ефект при використанні даної розробки. Аналізуючи розрахунки можна зробити висновок, що розроблена інформаційна технологія передбачення електрогенерації ФЕС з урахуванням впливу хмарності, за ціною дешевша за аналог і є конкурентоспроможною. Період окупності складе близько 1,01 рік. Отже, з урахуванням отриманих результатів можна зробити висновок про доцільність науково-технічної роботи у рамках цієї теми. Успішна комерціалізація може призвести до значних успіхів на ринку та забезпечити інвестора високим рівнем чистого прибутку у найкоротший термін.

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

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. До уваги громадськості та експертів - проект Дорожньої карти розвитку відновлюваної енергетики України на період до 2020 року. Державне агентство з енергоефективності та енергозбереження України. URL: <http://sae.gov.ua/uk/pressroom/1133>. 11.2023
2. Chowdhury A. A. Reliability Modeling of Distributed Generation in Conventional Distribution Systems Planning and Analysis. URL: <https://www.semanticscholar.org/paper/Reliability-modeling-of-distributed-generation-in-Chowdhury-Agarwal/577ab47d67714fc79069597f20cc2e4811b15661>. 11.2023
3. Bae I., Kim J. Reliability Evaluation of Distributed Generation Based on Operation Mode. URL: <https://ieeexplore.ieee.org/document/4162576>. 11.2023
4. Medeiros R., Xu X., Makram E. Assessment of Operating Condition Dependent Reliability Indices in Microgrids. Journal of Power and Energy Engineering. URL: <https://www.scirp.org/journal/paperinformation?paperid=66198>. 11.2023
5. Kondo T, Jumpei Baba, Akihiko Yokoyama. Voltage control of distribution network with a large penetration of photovoltaic generations using facts devices. URL: https://www.researchgate.net/publication/229904659_Voltage_Control_of_Distribution_Network_with_a_Large_Penetration_of_Photovoltaic_Generations_using_FACTS_Devices. 11.2023
6. 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
7. 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
8. Форкун Я. Б., Шкурпела О.О. Сонячна теплоенергетика, конспект лекцій. Харків: ХНУМГ ім.О.М.Бекетова, 2020. 63 с.

9. Кудря С.О., Нетрадиційні та відновлювані джерела енергії. Київ: Національний технічний університет України («КПІ»), 2012. 240 с.

10. Шклярю В.И., Дубровская В.В., Карпенко Д.С. Використання енергії сонця фотоелектричними системами. Київ: Національний технічний університет України («КПІ»), 2009. 218 с.

11. 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.

12. Guide on forecast models all around the world. URL: <https://windy.app/blog/what-is-a-weather-forecast-model-guide-on-forecast-models-all-around-the-world.html>. 11.2023

13. 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

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

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

16. Wang K., Qi X., Liu H. 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

17. Wu C., Zhang X.P., Sterling M. Solar power generation intermittency and aggregation. URL: <https://www.nature.com/articles/s41598-022-05247-2>. 11.2023

18. Järvelä M., Lappalainen K., Valkealahti S. Characteristics of the cloud enhancement phenomenon and PV power plants. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X19311909>. 11.2023

19. Zhang J., Verschae R., Nobuhara J.F. Deep photovoltaic nowcasting. URL: <https://www.scinapse.io/papers/2963986342>. 11.2023

20. Ryu A., Ito M., Ishii H. Preliminary analysis of short-term solar irradiance forecasting by using total-sky imager and convolutional neural network. URL: <https://ieeexplore.ieee.org/document/8715984>. 11.2023
21. AlSkaif T., Dev S., Visser L. A systematic analysis of meteorological variables for PV output power estimation. URL: https://www.researchgate.net/publication/339065072_A_systematic_analysis_of_meteorological_variables_for_PV_output_power_estimation. 11.2023
22. Huang H., Xu J., Peng Z. Cloud motion estimation for short term solar irradiation prediction. URL: <https://ieeexplore.ieee.org/document/6688040>. 11.2023
23. Zou K.H., Tuncali K., Silverman S.G. Correlation and simple linear regression. URL: <https://sites.utexas.edu/sos/guided/inferential/numeric/bivariate/cor/>. 11.2023
24. Shrestha A., Mahmood A. Review of deep learning algorithms and architectures. URL: <https://ieeexplore.ieee.org/document/8694781>. 11.2023
25. Ahmed R., Sreeram V., Mishra Y. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. URL: <https://www.sciencedirect.com/science/article/abs/pii/S1364032120300885>. 11.2023
26. Li Y., Ye F., Liu Z. A short-term photovoltaic power generation forecast method based on LSTM. URL: <https://www.hindawi.com/journals/mpe/2021/6613123/>. 11.2023
27. Nie Y., Li X., Paletta Q. Open-Source Ground-based Sky Image Datasets for Very Short-term Solar Forecasting, Cloud Analysis and Modeling. URL: https://www.researchgate.net/publication/365821657_Open-Source_Ground-based_Sky_Image_Datasets_for_Very_Short-term_Solar_Forecasting_Cloud_Analysis_and_Modeling_A_Comprehensive_Survey. 11.2023
28. Chris Hoffman A. Python Quick Reference / A. Chris Hoffman., 1995. URL: <https://www.python.org/ftp/python/doc/quick-ref.1.3.html>. 11.2023
29. Бібліотека Pandas. URL: <https://uk.wikipedia.org/wiki/Pandas>. 11.2023
30. Python. URL: <https://aws.amazon.com/ru/what-is/python/>. 11.2023

31. Data Science. URL: <https://coursehunter.net/course/data-science-uroven1-instrumenty-i-tehnologii>. 11.2023
32. Bandyopadhyay A. Hands-On GPU Computing with Python. URL: <https://www.packtpub.com/product/hands-on-gpu-computing-with-python/9781789341072>
33. Bruce E. Pease Leading Intelligence Analysis. CQ Press. URL: https://books.google.com.ua/books/about/Leading_Intelligence_Analysis.html?id=gqhkDwAAQBAJ&redir_esc=y
34. Blau D. M. A Time Series Analysis of Self Employment in the United States. URL: <https://www.jstor.org/stable/1831972>
35. Козловський В. О., Лесько О. Й., Кавецький В. В. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт: уклад. Вінниця: ВНТУ, 2021. 42 с.
36. Мокін В. Б. Heart Disease - Automatic AdvEDA & FE & 20 models. URL: <https://www.kaggle.com/vbmokin/heart-disease-comparison-of-20-models>
37. Weron R. Modeling and Forecasting Electricity Loads and Prices. URL: <http://prac.im.pwr.wroc.pl/~hugo/RePEc/wuu/hsbook/HSBook0601.pdf>

Додаток А

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 2. АНАЛІЗ ВПЛИВУ ХМАРНОСТІ

08-34.МКР.002.02.000 ТЗ

Керівник: д.т.н., проф. каф. САІТ

_____ Олександр МОКІН

«__» _____ 2023 р.

Розробив: студент гр. 2ІСТ-22м

_____ Михайло ГНАТЮК

«__» _____ 2023 р.

Вінниця 2023

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 „Energy” <https://www.kaggle.com/datasets/hnatyukmu/energy>

3. Мета і призначення роботи:

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

4. Вихідні дані для проведення робіт:

Набір даних моніторингу електрогенерації приватної сонячної електростанції розташованої на території Вінницької 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 електрогенерацію фотоелектростанцією, за доведеними даними: числом та місяцем року, прогнозом погоди (хмарності).

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 те-хнології» (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 _____ Михайло ГНАТЮК

Додаток Б

Протокол перевірки магістерської кваліфікаційної роботи на наявність
текстових запозичень

Назва роботи: «Аналіз та передбачення добового генерування електроенергії
фотоелектростанцією. Частина 2. Аналіз впливу хмарності»

Тип роботи: магістерська кваліфікаційна робота

Підрозділ: кафедра САІТ

Показники звіту подібності Unicheck

Оригінальність 93,3 %

Схожість 6,7 %

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)

Михайло ГНАТЮК

Керівник роботи



(підпис)

Олександр МОКІН

Додаток В

Лістинг програми

```
# Prediction of target for training data for all models
shift = 2000
count = 100
index = X_train.index[shift:shift+count]
X_train = X_train.loc[index]
y_train = y_train.loc[index]

y_pred_dtr = dtr.predict(X_train)
y_pred_rfr = rfr.predict(X_train)
y_pred_lgb = model_lgb.predict(X_train)

x = X_train["TOTAL_TIME"][index];

# Building plot for prediction for the training data
plt.figure(figsize=(20,15))
plt.scatter(x, y_train, label = "Target training data", color = 'g', marker='^')
plt.scatter(x, y_pred_dtr, label = "Decision Tree Regressor prediction", color = 'b',
marker='x')
plt.scatter(x, y_pred_rfr, label = "Random Forest prediction", color = 'y', marker='s')
plt.scatter(x, y_pred_lgb, label = "LightGBM Regressor prediction", color = 'brown',
marker='o')
plt.plot(x, np.full(len(X_train), 1200), label = "Maximum allowable value", color = 'r')
plt.title('Prediction for the training data')
plt.legend(loc='upper right')
plt.grid(True)
```

```
# Prediction of target for valid data for all models
shift = 1000
count = 100
index = X_valid.index[shift:shift+count]
X_valid = X_valid.loc[index]
y_valid = y_valid.loc[index]

y_val_dtr = dtr.predict(X_valid)
y_val_rfr = rfr.predict(X_valid)
y_val_lgb = model_lgb.predict(X_valid)

x = X_valid["TOTAL_TIME"][index];

# Building plot for prediction for the valid data
plt.figure(figsize=(20,15))
plt.scatter(x, y_valid, label = "Target valid data", color = 'g', marker='^')
plt.scatter(x, y_val_dtr, label = "Decision Tree Regressor prediction", color = 'b',
marker='x')
plt.scatter(x, y_val_rfr, label = "Random Forest prediction", color = 'y', marker='s')
plt.scatter(x, y_val_lgb, label = "LightGBM Regressor prediction", color = 'brown',
marker='o')
plt.plot(x, np.full(len(X_valid), 1200), label = "Maximum allowable value", color = 'r')
plt.title('Prediction for the valid data')
plt.legend(loc='upper right')
plt.grid(True)

# Prediction of target for test data for all models
shift = 2000
count = 100
```

```
index = X_test.index[shift:shift+count]
X_test = X_test.loc[index]
y_test = y_test.loc[index]

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

x = X_test["TOTAL_TIME"][index];

# Building plot for prediction for the test data
plt.figure(figsize=(20,15))
plt.scatter(x, y_test, label = "Target test data", color = 'g', marker='^')
plt.scatter(x, y_test_dtr, label = "Decision Tree Regressor prediction", color = 'b',
marker='x')
plt.scatter(x, y_test_rfr, label = "Random Forest prediction", color = 'y', marker='s')
plt.scatter(x, y_test_lgb, label = "LightGBM Regressor prediction", color = 'brown',
marker='o')
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)
```

Додаток Г

ІЛЮСТРАТИВНА ЧАСТИНА

АНАЛІЗ ТА ПЕРЕДБАЧЕННЯ ДОБОВОГО ГЕНЕРУВАННЯ ЕЛЕКТРОЕНЕРГІЇ
ФОТОЕЛЕКТРОСТАНЦІЄЮ. ЧАСТИНА 2. АНАЛІЗ ВПЛИВУ ХМАРНОСТІ

Нормоконтроль: к.т.н., доцент

_____Сергій ЖУКОВ

«__» _____ 2023 р.

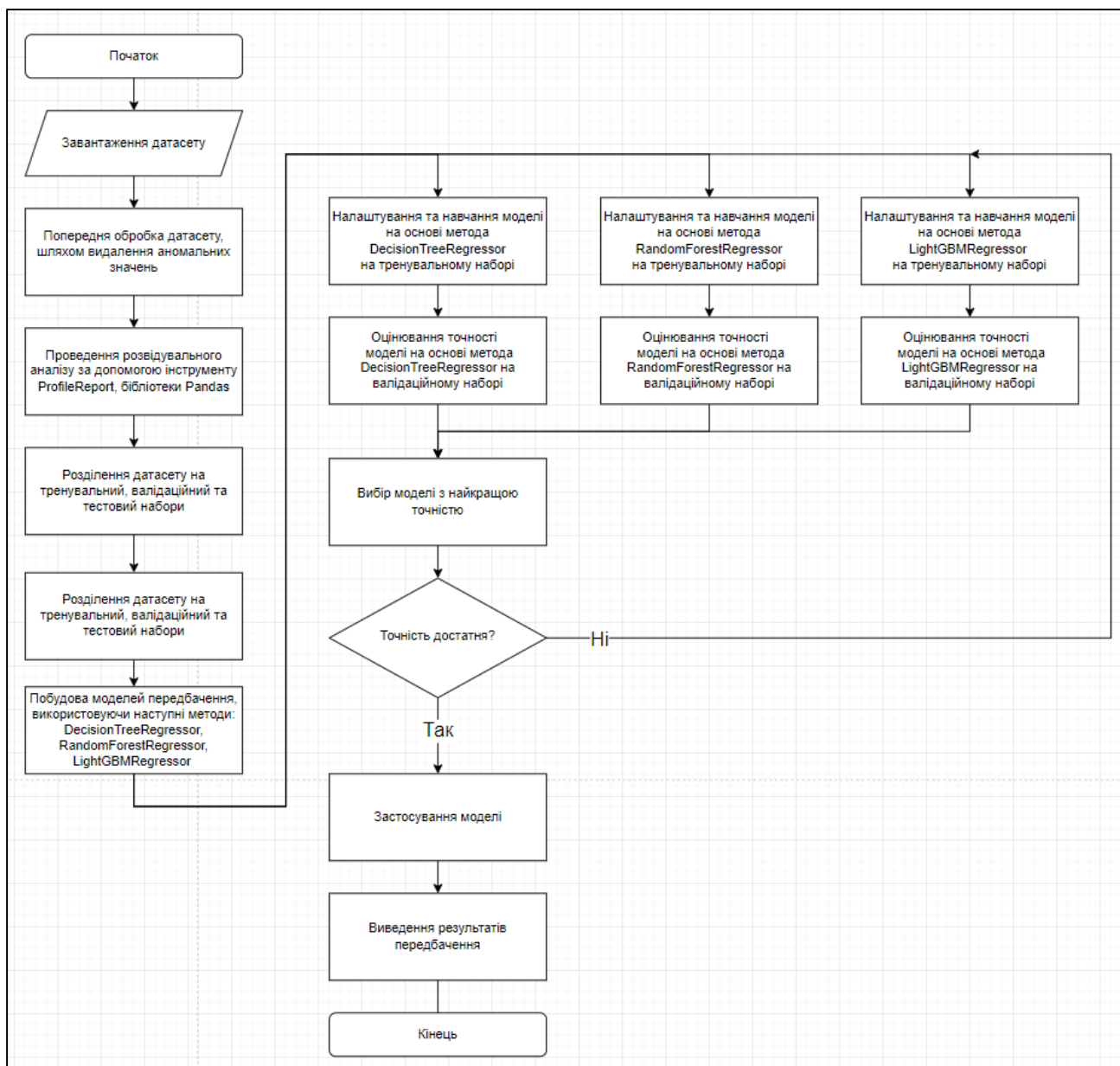


Рисунок Г.1 – Блок-схема алгоритму інформаційної технології передбачення електрогенерації фотоелектростанції

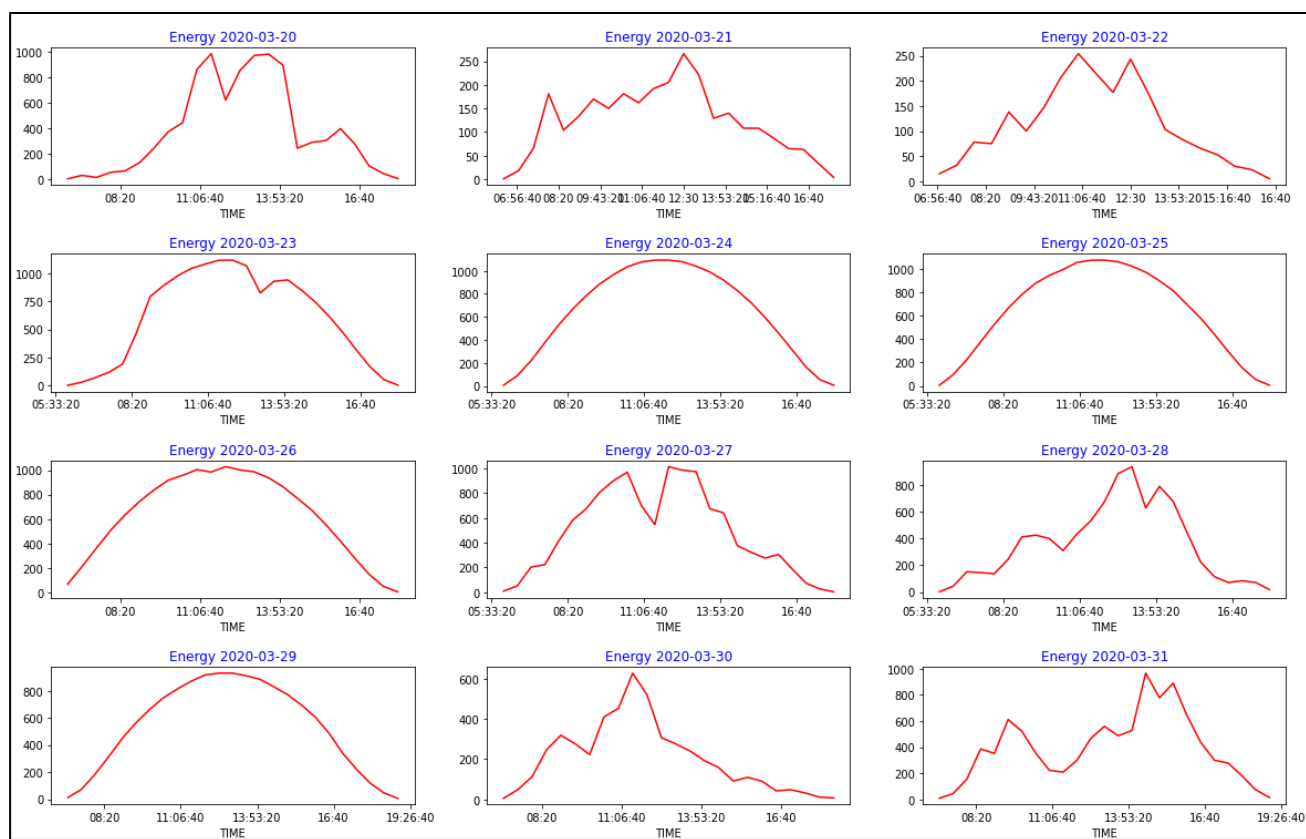


Рисунок Г.2 – Добові графіки генерації електроенергії

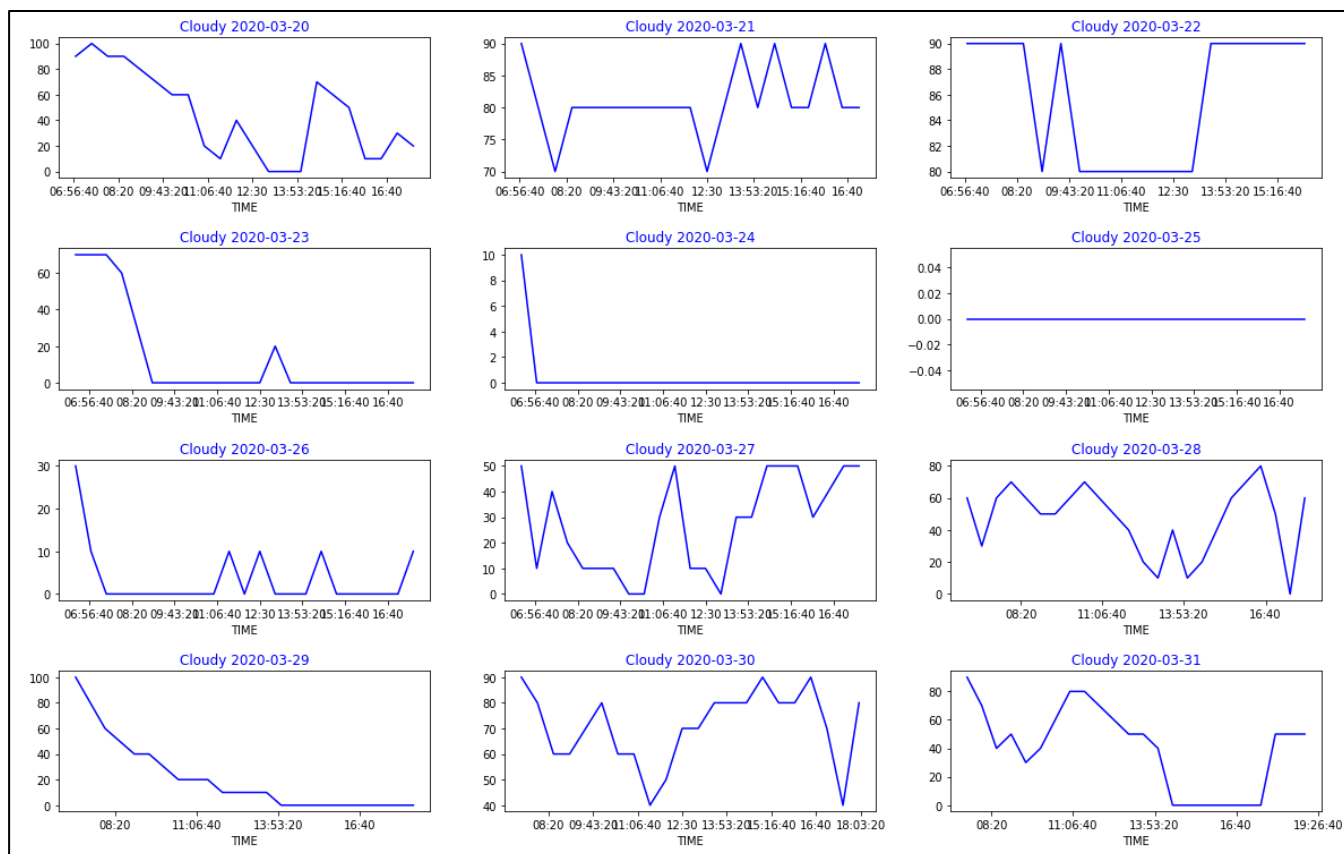


Рисунок Г.3 – Добові графіки зміни хмарності

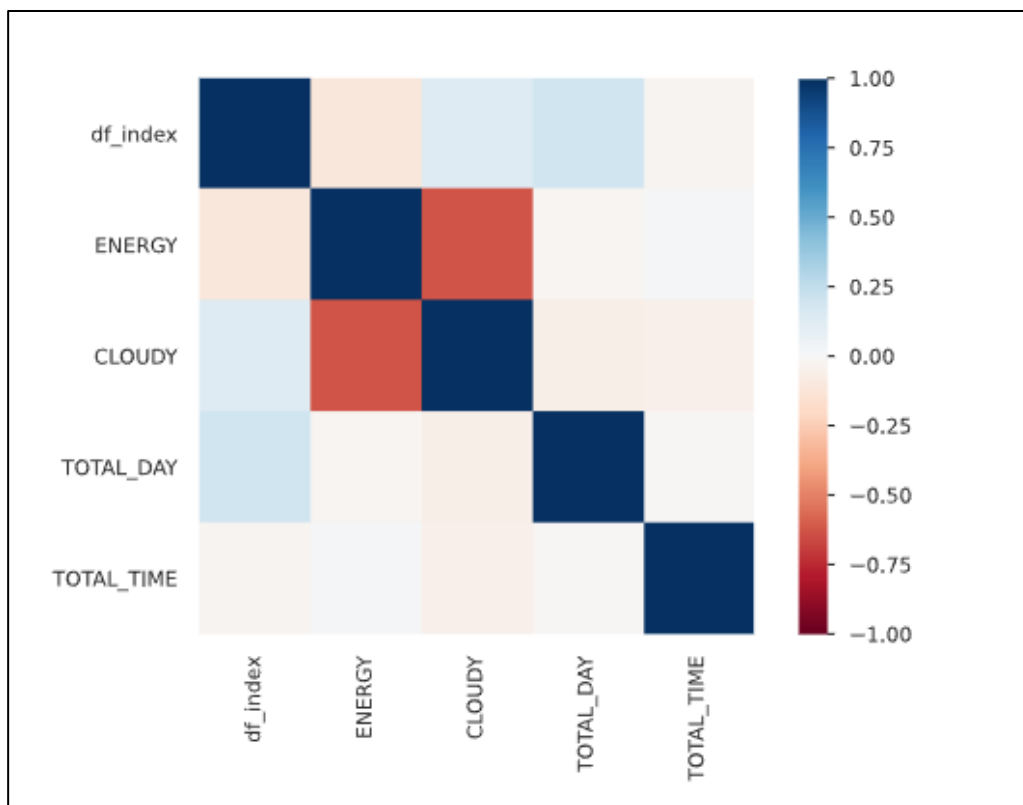


Рисунок Г.4 – Матриця кореляції Пірсона

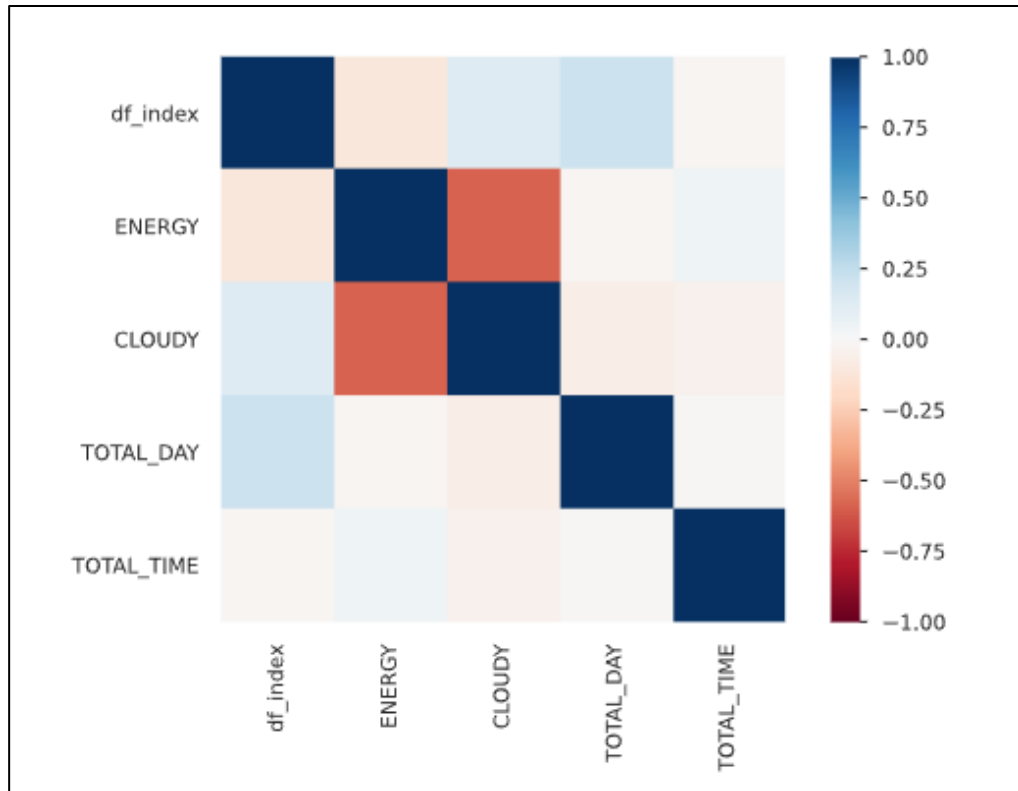


Рисунок Г.5 – Матриця кореляції Спірмена

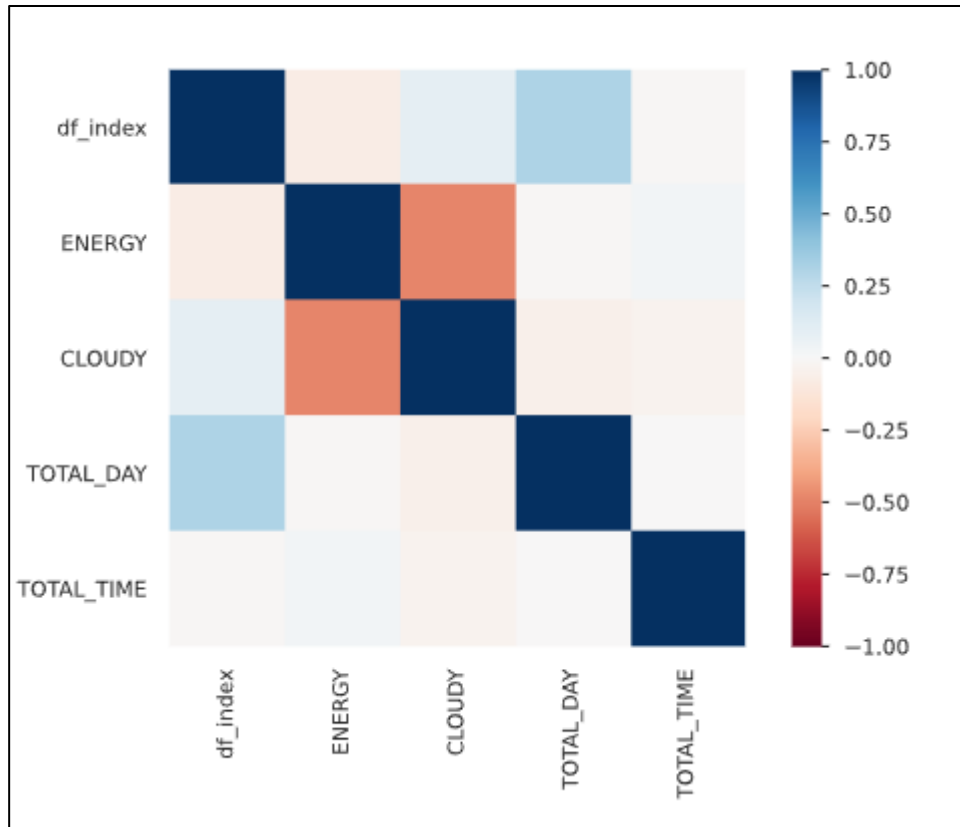


Рисунок Г.6 – Матриця кореляції Кендалла

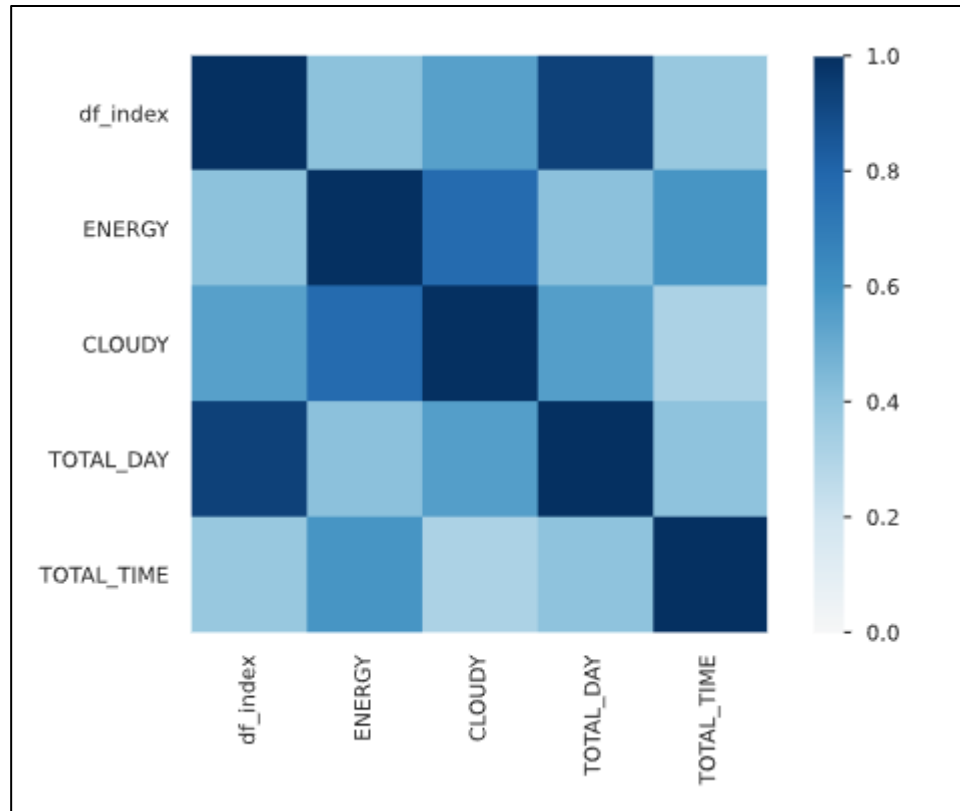


Рисунок Г.7 – Матриця кореляції Rho_{ik}

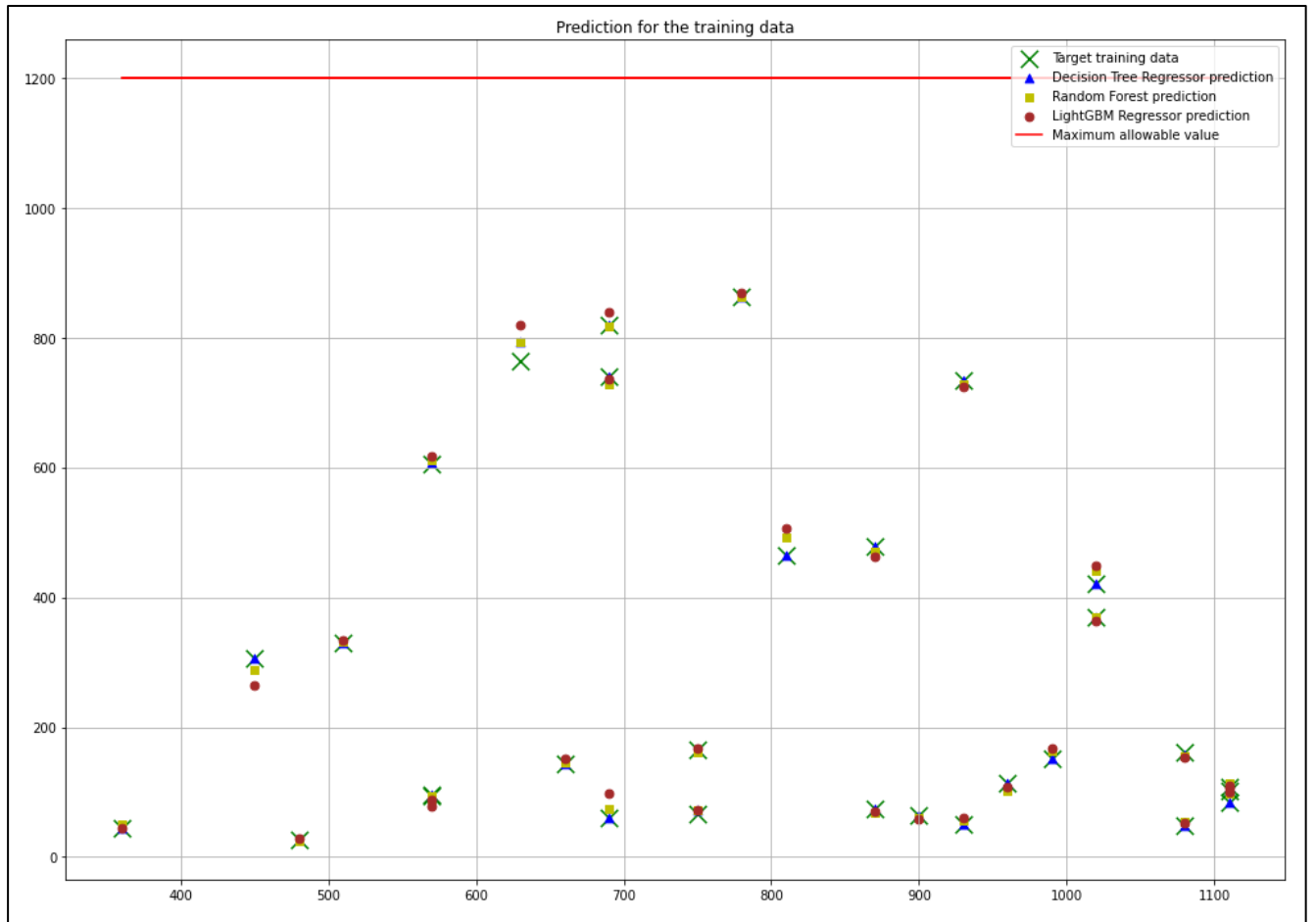


Рисунок Г.8 – Передбачення навчальних даних

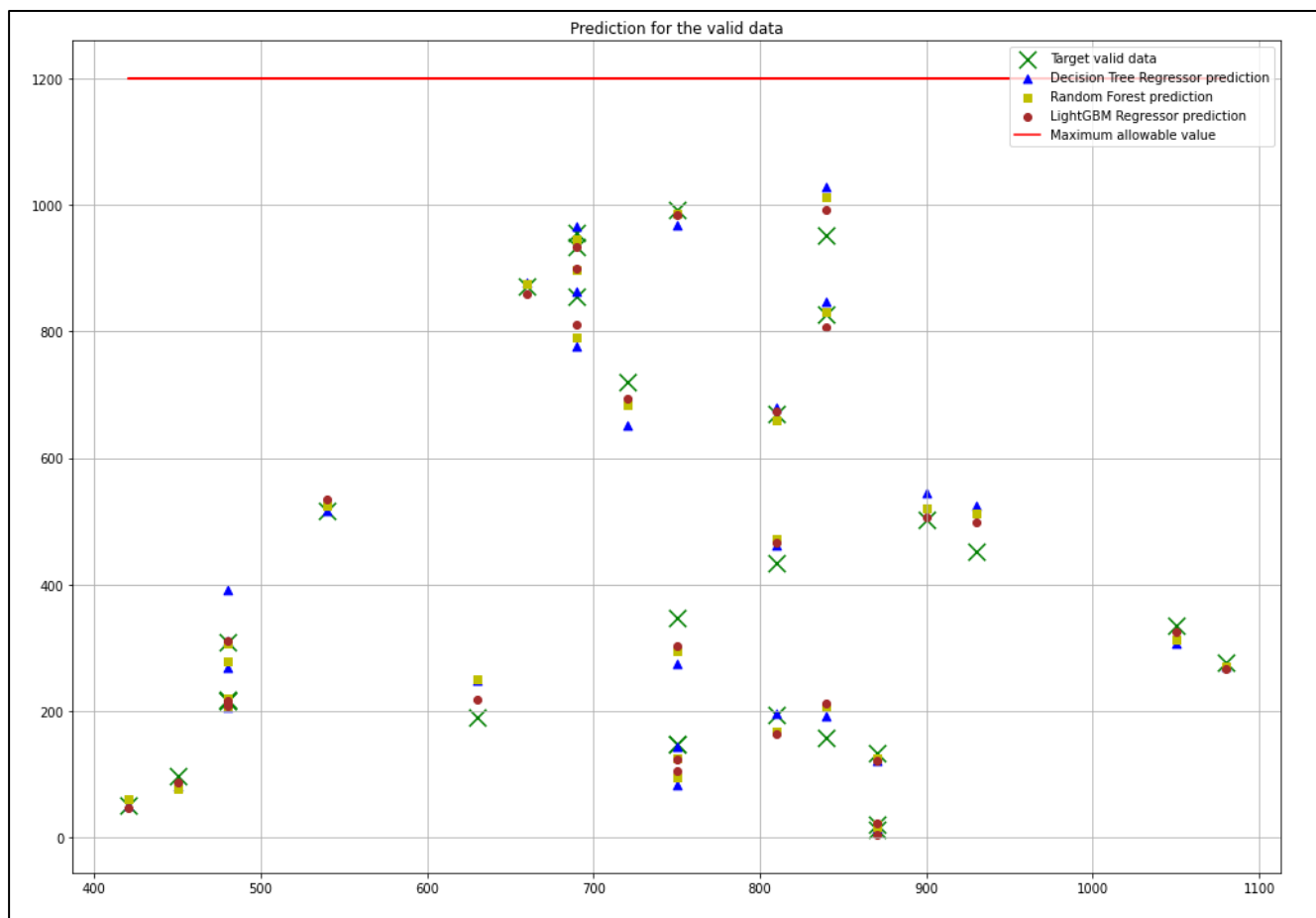


Рисунок Г.9 – Передбачення валідаційних даних

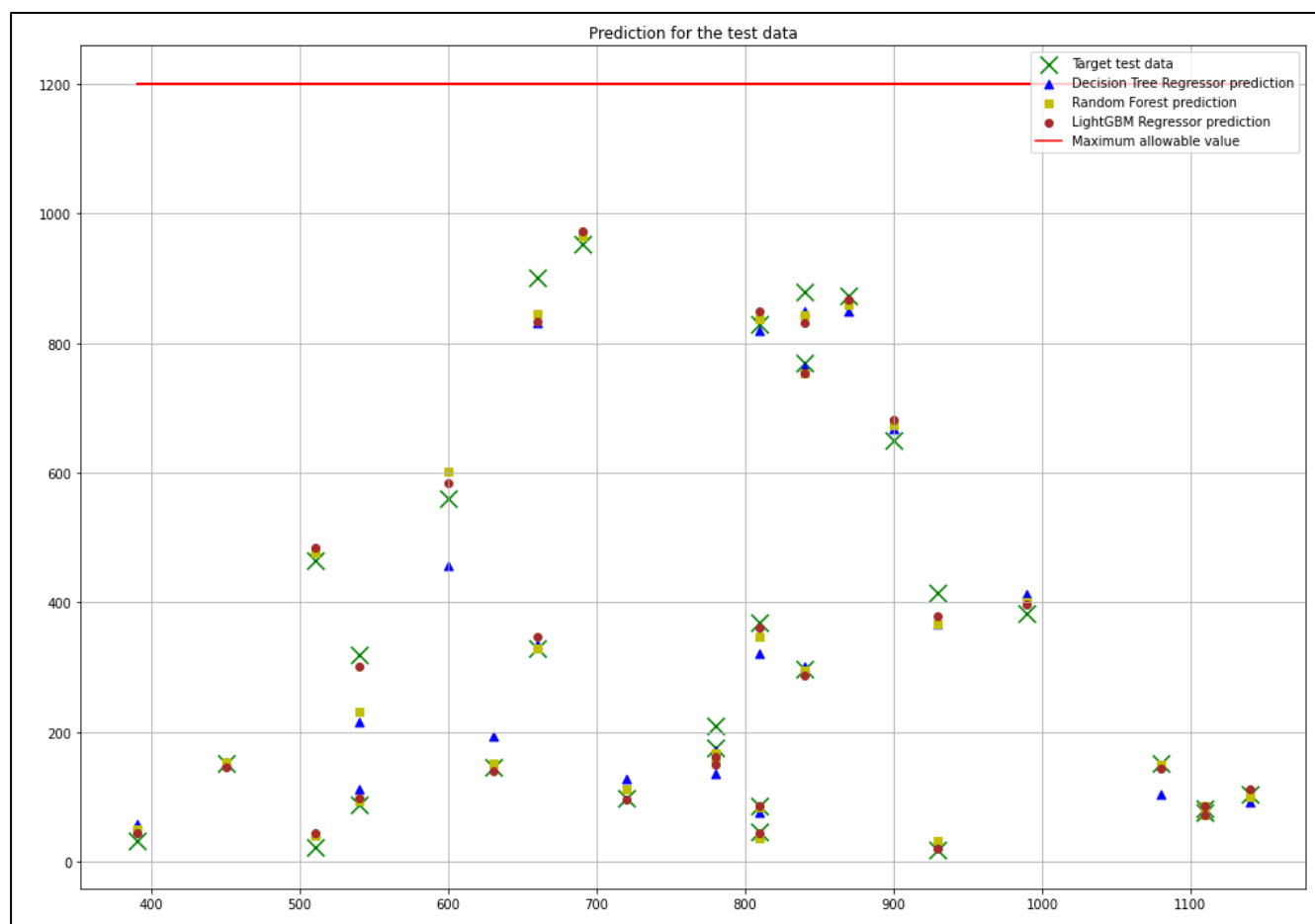


Рисунок Г.10 – Передбачення тестових даних

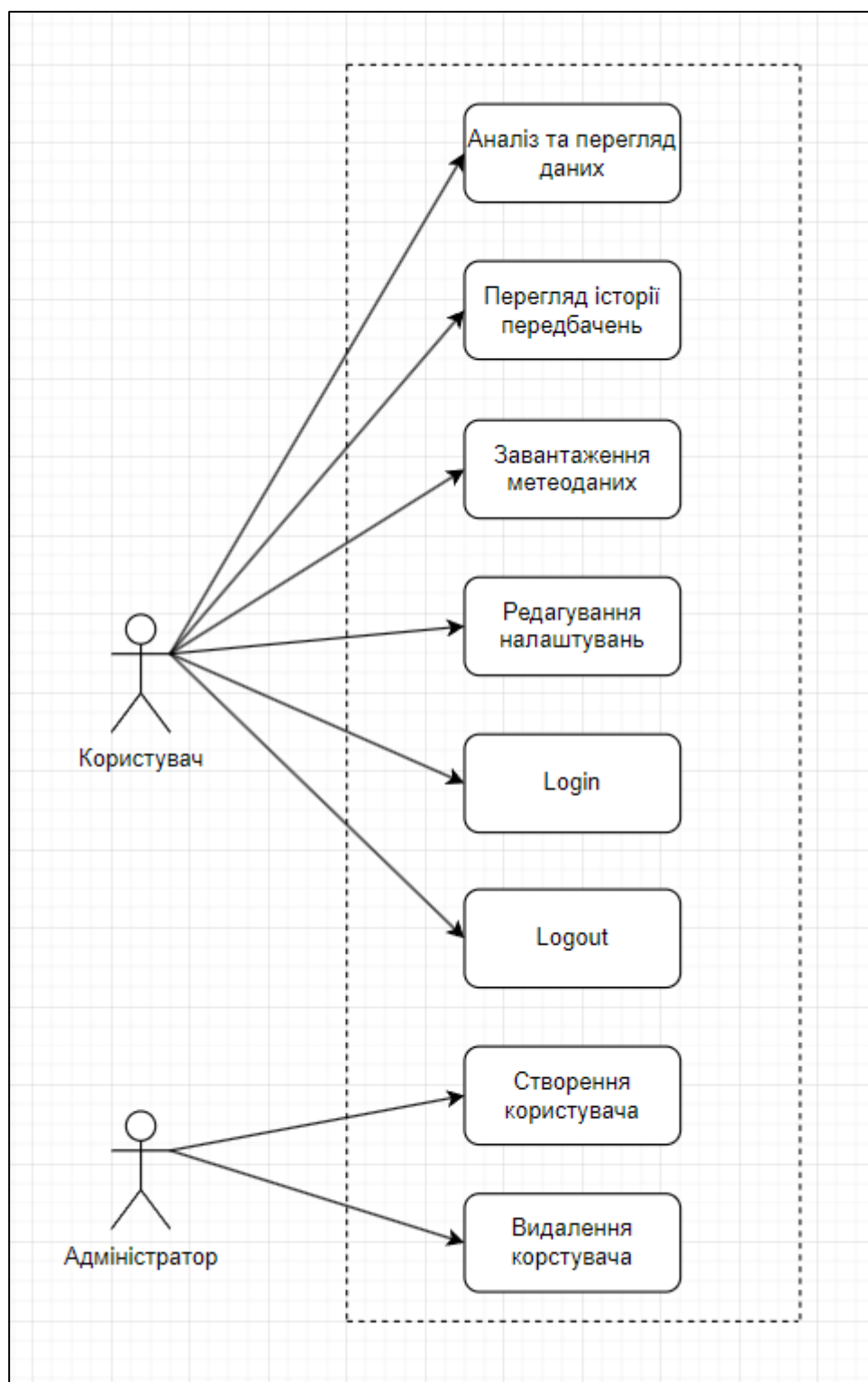


Рисунок Г.11 – Use-Case діаграма розробленої інформаційної технології