

DOI: [https://doi.org/10.25140/2411-5363-2026-2\(44\)-584-599](https://doi.org/10.25140/2411-5363-2026-2(44)-584-599)

УДК 004.89; 621.3; 351/354; 658

**Олександр Миколайович Васильєв<sup>1</sup>, Сергій Сергійович Зубенко<sup>2</sup>**

<sup>1</sup>аспірант 2 року навчання, спеціальність 193 – Геодезія та землеустрій  
Київський національний університет імені Тараса Шевченка ННІ «Інститут геології» (Київ, Україна)

E-mail: [yasilyeva455@gmail.com](mailto:yasilyeva455@gmail.com). ORCID: <https://orcid.org/0009-0009-2177-7859>

<sup>2</sup>магістр 2 року навчання, спеціальність 103 – Науки про Землю.  
Київський національний університет імені Тараса Шевченка ННІ «Інститут геології» (Київ, Україна)

E-mail: [szubnko02@gmail.com](mailto:szubnko02@gmail.com). ORCID: <https://orcid.org/0009-0008-1166-4693>

**ТЕХНОЛОГІЇ BIG DATA В ПРОГНОЗУВАННІ ГЕНЕРАЦІЇ СОНЯЧНОЇ ЕНЕРГІЇ**

*У дослідженні розглянуто можливості застосування технологій Big Data для прогнозування генерації сонячної електроенергії. Проаналізовано особливості часових рядів виробництва електроенергії з відновлюваних джерел та вплив природно-кліматичних чинників на змінність генерації. Для дослідження використано відкритий набір даних Open Power System Data, що містить погодинні показники виробництва енергії. Проведено попередню обробку та статистичний аналіз даних, виконано декомпозицію часових рядів і побудовано алгоритм прогнозування. Отримані результати підтверджують ефективність використання методів аналізу великих масивів даних для підвищення точності прогнозування генерації сонячної енергії та підтримки управлінських рішень у сфері енергетики.*

**Ключові слова:** великі дані, Big Data, оптимізація виробництва сонячної енергії, прогнозування.

Табл.: 1. Рис.: 14. Бібл.: 20.

**Актуальність теми дослідження.** Розвиток відновлюваної енергетики є одним із ключових напрямів забезпечення сталого розвитку та енергетичної безпеки. Водночас сонячна генерація характеризується високою часовою, добовою, сезонною мінливістю, що ускладнює процеси прогнозування й управління енергосистемами. Існуючі методи прогнозування мають певні обмеження і не завжди забезпечують достатню точність особливо в умовах динамічних змін навколишнього середовища і його погодних умов. Тому актуальним науково-практичним завданням є розробка нових підходів, заснованих на технологіях великих даних для підвищення ефективності прогнозування та оптимізації виробництва сонячної енергії, дослідження часових закономірностей виробництва електроенергії з ВДЕ.

Зростання попиту на ВДЕ та інтеграція сонячної енергетики в сучасні енергетичні системи зумовлюють потребу в точному прогнозуванні генерації електроенергії для забезпечення надійності та стабільності мережі. Точні прогнози сонячної генерації дозволяють ефективніше балансувати попит і пропозицію, зменшувати витрати на резервні потужності та підвищувати загальну енергоефективність енергосистеми. Попередні дослідження засвідчили, що використання машинного та глибокого навчання, метеорологічних даних для аналізу часових рядів генерації дозволяє зменшити небаланси та підвищити економічну ефективність роботи систем, сприяє підвищенню точності коротко- та довгострокових прогнозів виробництва сонячної електроенергії.

Сонячні електростанції характеризуються високою змінністю виробництва, що залежить від погодних умов, часу доби та сезонних коливань сонячної радіації. Ця мінливість ускладнює планування режимів роботи енергосистеми та потребує впровадження ефективних методів прогнозування генерації для зменшення небалансів.

Використання високоточних метеопрогнозів, супутникових даних і алгоритмів машинного навчання дозволяє підвищити надійність оцінок виробітку сонячних електростанцій у коротко- та довгостроковій перспективі.

Сонячні електростанції характеризуються високою змінністю виробництва, що залежить від погодних умов, часу доби та сезонних коливань сонячної радіації. Недостатньо точні прогнози генерації ускладнюють балансування енергосистеми та підвищують ризики нестачі або надлишку електроенергії.

Застосування технологій Big Data дозволяє обробляти великі обсяги даних з різних джерел – метеорологічні показники, дані супутникового моніторингу, історичні дані ро-

боти панелей – та використовувати їх для створення високоточних моделей прогнозування. Це підвищує ефективність планування виробництва, оптимізує роботу енергосистеми та сприяє інтеграції відновлюваних джерел енергії у національну мережу.

Таким чином, дослідження методів аналізу великих даних для прогнозування генерації сонячної енергії є актуальним з погляду забезпечення енергетичної безпеки, економічної ефективності та сталого розвитку енергетики.

**Дані та методи дослідження.** У дослідженні використано відкриті набори даних платформи Open Power System Data (Німеччина), що містять детальні часові ряди виробництва електроенергії з відновлюваних джерел, зокрема сонячної та вітрової генерації. Дані представлені з високою часовою розрізненістю (годинні та субгодинні інтервали), що дозволяє виконувати аналіз короткострокових коливань і сезонних закономірностей генерації. Використані дані базуються на офіційній статистиці операторів енергосистем і ринку електроенергії Німеччини, що забезпечує їхню достовірність, повноту та придатність для наукових досліджень.

Методологія дослідження включає статистичний аналіз, декомпозицію часових рядів і виявлення трендових та сезонних компонент. Окремо аналізується вплив метеорологічних факторів на коливання обсягів генерації.

**Аналіз сучасних досліджень і публікацій.** За останні роки інтерес до застосування технологій Big Data у сфері відновлюваної енергетики, зокрема сонячної, значно зріс. Велика кількість досліджень присвячена методам збору, обробки та аналізу великих обсягів даних для точного прогнозування генерації електроенергії та оптимізації роботи енергетичних систем.

Підходи машинного навчання та штучного інтелекту для прогнозування генерації сонячної енергії представлені в працях [1, 2] і свідчать, що використання алгоритмів нейронних мереж, регресійного аналізу та методів ансамблевого навчання дозволяє підвищити точність прогнозів у порівнянні з традиційними статистичними методами.

Дослідження представлені в працях [3, 4] демонструють ефективність інтеграції даних супутникового спостереження та метеорологічних показників для побудови моделей прогнозування, що враховують сезонні та добові коливання сонячної радіації. Такі підходи дозволяють отримувати прогнози з короткостроковою (години) та середньостроковою (дні) точністю, що є критично важливим для управління енергосистемою.

У працях [5, 6] підкреслюється важливість великих даних та технологій IoT для моніторингу стану сонячних панелей у реальному часі. Це дозволяє відстежувати деградацію модулів, втрати ефективності та адаптувати прогнозні моделі відповідно до фактичних умов.

Водночас сучасні дослідження [7] вказують на наявні обмеження: потребу у великих обчислювальних потужностях, складність інтеграції даних з різних джерел та необхідність корекції моделей під локальні кліматичні умови.

Таким чином, аналіз наукових публікацій показує, що технології Big Data є перспективним інструментом для прогнозування генерації сонячної енергії, проте потребують комплексного підходу, що включає обробку великих обсягів даних, застосування сучасних алгоритмів машинного навчання та адаптацію моделей до специфіки конкретної енергетичної системи.

У праці [8] досліджено ринкові механізми стимулювання розвитку систем накопичення енергії в Україні, зосереджено увагу на аналізі регуляторного середовища, моделей державної підтримки та економічних інструментів, що можуть забезпечити інтеграцію накопичувачів енергії в енергетичний ринок. Розглянуто можливості впровадження стимулюючих тарифів, участі систем накопичення в ринку допоміжних послуг, а також перспективи створення конкурентних умов для інвестування в технології зберігання електроенергії.

Можливості підвищення енергонезалежності територіальних громад, розглянуто в праці [9]. Питання енергоефективності та енергозалежності в Україні розглянуто в праці [10]. Перетворення енергії різних видів відновлюваних джерел – вітрової, сонячної та геотермальної енергії, енергії малих річок і біомаси – на електричну й теплову енергію розглянуто в праці [11].

Огляд методів машинного навчання для прогнозування фотоелектричної енергії представлено в праці [12]. В праці [13] показано, що попри швидке зростання енергії з ВДЕ, його мінливий характер ускладнює його прогнозування.

Таким чином, проведений аналіз літературних джерел свідчить, що проблема ефективного використання сонячної енергетики і ВДЕ загалом, їх прогнозування шляхом застосування сучасних технологій аналізу великих даних є вкрай актуальною задачею.

**Мета дослідження** – оцінка ефективності використання технологій Big Data для підвищення точності прогнозування виробництва сонячної енергії та обґрунтування їх ролі в забезпеченні стабільності й адаптивності сучасних енергетичних систем.

**Наукова новизна** дослідження полягає у комплексному підході до інтеграції технічних, метеорологічних і споживчих даних у межах єдиного аналітичного середовища big data для прогнозування генерації сонячної енергії, що дає змогу підвищити точність прогнозів та обґрунтувати управлінські рішення в енергетичному секторі.

**Завдання дослідження:** аналіз можливості застосування Big Data в енергетичному секторі; інформаційної інфраструктури збору та збереження енергетичних даних; реалізація аналізу даних OPSD Germany в середовищі Jupyter Notebook.

**Виклад основного матеріалу.** Технології Big Data відіграють ключову роль у сьогоднішній трансформації енергетичного сектору, зокрема у сфері альтернативної енергетики, де виникає потреба в аналізі величезних обсягів гетерогенних даних із різних джерел. Це пов'язано з тим, що специфіка надання послуг енергозабезпечення, обсягів споживання та виробництва енергії, стану обладнання, режимів роботи ліній електропередачі, прогнозування погодних умов тощо передбачає необхідність обробки великих масивів даних в режимі реального часу.

Альтернативна енергетика охоплює системи генерації з відновлюваних джерел – сонячної, вітрової, гідро-, геотермальної та біоенергетики. Ці системи характеризуються високою варіабельністю виробництва, залежністю від погодних і кліматичних умов, а також необхідністю збалансування між генерацією, зберіганням та споживанням енергії в реальному часі. Тому саме технології Big Data стають центральним елементом аналітичної інфраструктури, здатним забезпечити інтеграцію, обробку та прогнозування на основі багатовимірних і високочастотних даних.

Серед напрямків реального застосування технологій Big Data в енергетиці можна відзначити: взаємодію з клієнтами, управління та інтелектуалізацію мереж, створення мікромереж, віртуальних станцій, нових бізнес-моделей на ринку, торгівля електроенергією, її акумулювання та виявлення крадіжок, прогнозування режимів роботи інфраструктури та планування її розвитку, підвищення рівня кібербезпеки енергосистем, управління виробничими процесами тощо.

Великі дані (Big Data) сьогодні відіграють ключову роль у розвитку енергетичних систем нового покоління, особливо у сфері ВДЕ. Їх використання дозволяє оптимізувати прогнозування генерації, управління навантаженнями, балансування мережі та зменшення втрат енергії.

Основними джерелами енергетичних даних виступають: міжнародні організації, такі як МЕА, ООН або Світовий банк, які надають узагальнені та порівняльні енергетичні дані між країнами; енергетичні компанії та оператори мереж, які надають дані про видо-

буток, генерацію, передачу і розподіл енергії; наукові установи та університети, які публікують аналітичні та експериментальні енергетичні дані у звітах і наукових статтях; відкриті онлайн-платформи та бази відкритих даних, які забезпечують доступ до енергетичної статистики в цифровому форматі; сенсорні мережі, метеостанції, супутникові системи дистанційного зондування Землі (ДЗЗ), смартлічильники, енергетичні біржі та хмарні платформи моніторингу тощо. Кожен із цих елементів генерує потоки інформації у різних форматах (часові ряди, геопросторові зображення, телеметричні сигнали), що потребують інтеграції в єдину базу даних для подальшої аналітики. Використання хмарних обчислень (AWS, Google Cloud, Azure) у поєднанні з розподіленими системами збереження, такими як Hadoop, Spark та Cassandra, дає змогу виконувати масштабні обчислення з терабайтними наборами даних. У контексті альтернативної енергетики це дозволяє створювати реалістичні цифрові моделі енергосистем, здатні реагувати на зміни зовнішніх умов у режимі реального часу.

Аналітична обробка великих енергетичних даних передбачає застосування методів машинного навчання (МН) і глибокого навчання (ГН) для виявлення закономірностей, прогнозування попиту та оптимізації роботи енергосистем. При цьому, методи МН використовуються головним чином для класифікації, кластеризації та регресійного аналізу енергетичних даних, а методи ГН особливо ефективні в роботі з великими обсягами часових рядів і даних з датчиків у «розумних» мережах та системах енергомоніторингу. Завдяки цьому підвищується точність прогнозів, надійність енергопостачання та енергоефективність.

Моделі прогнозування попиту й генерації енергії будуються з урахуванням сезонних і стохастичних компонентів часових рядів, оскільки споживання та виробництво енергії мають регулярні періодичні коливання і випадкові збурення. Сезонні компоненти відображають добові, тижневі та річні цикли, пов'язані з поведінкою споживачів і погодними умовами. Стохастичні компоненти враховують випадкові відхилення, похибки вимірювань та непередбачувані події в енергосистемі. Поєднання цих компонентів у моделях, таких як ARIMA, SARIMA, Prophet або нейронні мережі, забезпечує більш точні й надійні прогнози.

Для короткострокового прогнозування часових рядів широко застосовуються статистичні методи (ARIMA, SARIMA, Prophet) та нейронні підходи (RNN, LSTM), які дозволяють моделювати тренди, сезонність і нелінійні залежності [14].

Водночас для оцінювання ефективності роботи окремих елементів енергосистем, наприклад, прогнозування потужності вітрових турбін або сонячних панелей широко застосовуються регресійні моделі, ансамблеві методи машинного навчання (зокрема Random Forest і Gradient Boosting), а також штучні нейронні мережі [15].

Просторовий аналіз і візуалізація результатів досліджень здійснюються з використанням геоінформаційних платформ, які забезпечують інтеграцію енергетичних даних із географічними характеристиками, комплексно враховувати кліматичні та соціально-економічні чинники, підвищуючи обґрунтованість аналізу та підтримку управлінських рішень у сфері енергетики [15].

Ключовим чинником аналітики Big Data в енергетиці є концепція «data-driven decision making» – прийняття управлінських і технічних рішень на основі аналізу даних, а не лише інтуїції чи експертних припущень. У межах цієї концепції великі масиви енергетичних даних із лічильників, датчиків, SCADA та ринків електроенергії перетворюються на знання за допомогою аналітичних моделей. Data-driven підхід дає змогу оптимізувати прогнозування попиту, управління генерацією, зменшення втрат і підвищення надійності енергосистем. Таким чином, аналітика Big Data стає фундаментом цифрової трансформації сучасної енергетики.

Стратегічне та оперативне управління енергосистемами здійснюється не лише на основі експертних оцінок, але й із використанням моделей прогнозування, які безперервно оновлюються з урахуванням нових даних. Такий підхід забезпечує підвищення адаптивності, точності управлінських рішень і стійкості функціонування енергосистем в умовах невизначеності та змінного попиту.

У поєднанні з технологіями Інтернету речей (IoT) підходи Big Data формують передумови для розвитку розумних енергетичних мереж (Smart Grids), що забезпечують автоматизоване управління процесами генерації, споживання та акумулювання енергії. Інтеграція цих технологій підвищує ефективність, надійність і гнучкість енергосистем, сприяючи їх адаптації до динамічних змін попиту та зростання частки відновлюваних джерел енергії [16-18].

Застосування Big Data в альтернативній енергетиці також охоплюють питання надійності і якості даних, сумісності форматів, етичних аспектів і кібербезпеки, оскільки без надійних і захищених даних аналітичні рішення втрачають ефективність. Якість даних передбачає повноту, точність, актуальність і узгодженість інформації, що надходить з різних джерел відновлюваної енергетики. Сумісність форматів є критичною для інтеграції даних з різних платформ, виробників обладнання та інформаційних систем. Етичні аспекти стосуються конфіденційності даних споживачів і прозорості використання аналітичних моделей. Кібербезпека забезпечує захист енергетичних даних і критичної інфраструктури від несанкціонованого доступу та кібератак. Надійність моделей залежить від репрезентативності даних і стійкості алгоритмів до шуму. Тому розробка методів очищення, нормалізації та узгодження даних є критично важливою для достовірності результатів. Поєднання Big Data з технологіями штучного інтелекту створить умови для повноцінного переходу до енергетики п'ятого покоління – інтелектуальної, адаптивної й сталорозвиненої енергетичної екосистеми [19].

Для дослідження часових закономірностей виробництва електроенергії з відновлюваних джерел з урахуванням добової, сезонної та міжрічної мінливості була використана платформа OPSD (Open Power System Data), яка збирає, оброблює і публікує високоякісні централізовані дані про системи електропостачання Західної Європи з метою розробки моделей енергосистем, підвищення прозорості досліджень і скорочення дублювання роботи для дослідників у галузі енергетики за рахунок надання надійних вихідних даних.

OPSD розробляється за підтримки наукових інституцій та енергетичних аналітичних центрів і надає безкоштовний доступ до узгоджених даних про виробництво, споживання, потужності, ціни та викиди CO<sub>2</sub> у форматах CSV і API. Платформа агрегує та гармонізує інформацію з національних операторів електромереж, європейських агентств (ENTSO-E, Eurostat) і метеорологічних служб [20]. Дані охоплюють 30 країн Європи й містять часові ряди з інтервалом до 1 години починаючи з 2006 року. Особливістю OPSD є відкрите ліцензування, чітка структура метаданих і автоматизовані механізми перевірки якості. OPSD широко використовується в дослідженнях, пов'язаних із прогнозуванням виробництва енергії, моделюванням балансів електромереж, аналізом інтеграції відновлюваних джерел енергії та розробкою енергетичної політики. У поєднанні з аналітичними платформами (Python, Jupyter, Pandas) цей ресурс є базовим джерелом для реплікабельних наукових досліджень у сфері Big Data та альтернативної енергетики.

Дані Open Power System Data Germany характеризуються високою часовою розрізненістю та відкритістю, що робить їх придатними для статистичного аналізу, машинного навчання й побудови моделей прогнозування генерації. У дослідженні використано часові ряди за 2015–2020 рр., де основними ознаками виступають фактична генерація сонячної та вітрової енергії (у МВт).

Структура набору даних представлена в табл. 1.

Таблиця 1 – Структура набору даних

Назва поля	Тип даних	Одиниця виміру	Опис	Приклад значення
utc_timestamp	date-time	UTC (ISO 8601)	Часова мітка у форматі UTC, що визначає момент вимірювання енергогенерації. Використовується як уніфікований часовий індекс для об'єднання з іншими джерелами даних.	2015-01-01T00:00:00+00:00
cet_cest_timestamp	date-time	CET/CEST (UTC+1/UTC+2)	Локальний час (Central European Time або Central European Summer Time), що відповідає німецькому часовому поясу. Використовується для зручності аналізу добових циклів генерації.	2015-01-01T01:00:00+01:00
DE_solar_generation_actual	float	Мегавати (MW)	Фактична величина виробництва сонячної енергії у Німеччині, виміряна за годину. Відсутність значення (NaN) вказує на нічний період або відсутність даних.	773.0
DE_wind_generation_actual	float	Мегавати (MW)	Фактична величина виробництва вітрової енергії у Німеччині, виміряна за той самий період. Дані використовуються для моделювання короткострокового прогнозу генерації.	9689.00

Попередня обробка даних була спрямована на підвищення їхньої якості та придатності до подальшого аналізу. Зокрема, виконано очищення набору даних від пропусків, частка яких становила близько 0,3 % записів. Відсутні значення відновлювалися за даними попереднього дня, що дало змогу зберегти добову та сезонну структуру часових рядів. Крім того, здійснено усереднення добових показників з метою зниження рівня випадкового шуму для зниження шуму та підвищення стійкості подальших моделей аналізу і прогнозування. Фрагмент коду імпортування модулів представлено на рис. 1.

```

[2]: # Import packages needed for the project

import numpy as np
import pandas as pd
import itertools
import matplotlib.pyplot as plt
import seaborn as sns
from matplotlib.dates import DateFormatter
import datetime as dt
from statsmodels.tsa.stattools import adfuller
import plotly.offline as pyoff
import plotly.graph_objs as go
%matplotlib inline

import warnings
warnings.filterwarnings("ignore")

# matplotlib style & parameter changes
import matplotlib
matplotlib.style.use('seaborn-darkgrid')
matplotlib.rcParams['figure.figsize'] = (16, 7)

[3]: # Bring in all the data for 2015- 2021 Germany. DE is the two letter symbol for germany.
energy_all = pd.read_csv("time_series_60min_singleindex_filtered (3).csv",
                        parse_dates=[0], index_col=0)

[4]: # check out the info of the data frame
energy_all.info()

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 50401 entries, 2014-12-31 23:00:00+00:00 to 2020-09-30 23:00:00+00:00
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype

```

Рис. 1. Фрагмент коду імпортування модулів в Jupyter Notebook

Джерело: розроблено авторами.

Всього в датасеті нараховується 50401 рядок Тип даних визначено як float64, що забезпечує достатню числову точність для представлення значень генерації електроенергії та коректне виконання статистичного аналізу й побудови моделей прогнозування.

Проведений авторами статистичний свідчить про зростаючі тенденції виробництва німецьких відновних джерел енергії. Для подальшого аналізу потрібно позбутися стовпця `cet_cest_timestamp`. Перші значення набору даних представлені на рис. 2.

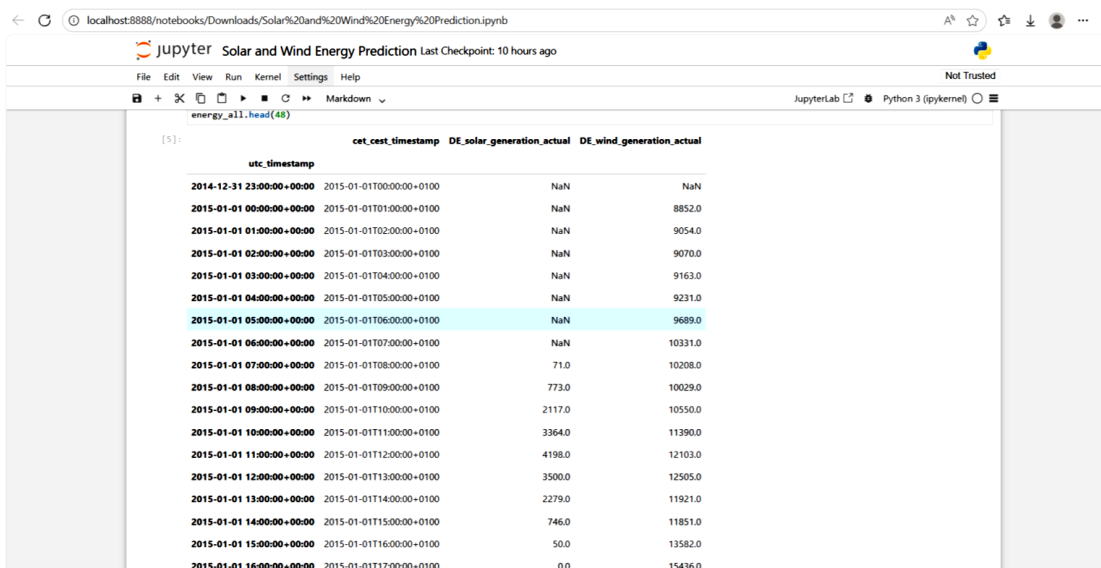


Рис. 2 Перші значення набору даних

Джерело: побудовано авторами в Jupyter Notebook.

Видаляти нульові значення недоцільно, оскільки це може порушити структуру та безперервність часового ряду

Значення NaN у часових рядах генерації сонячної та вітрової енергії відповідають відсутнім або некоректно зафіксованим спостереженням, що можуть виникати через збої виміррювального обладнання, затримки передачі даних або технічні помилки під час збору інформації.

Як свідчить рис. 3, є 7 точок даних NaN для сонячної енергії, і, крім того, багато нулів.

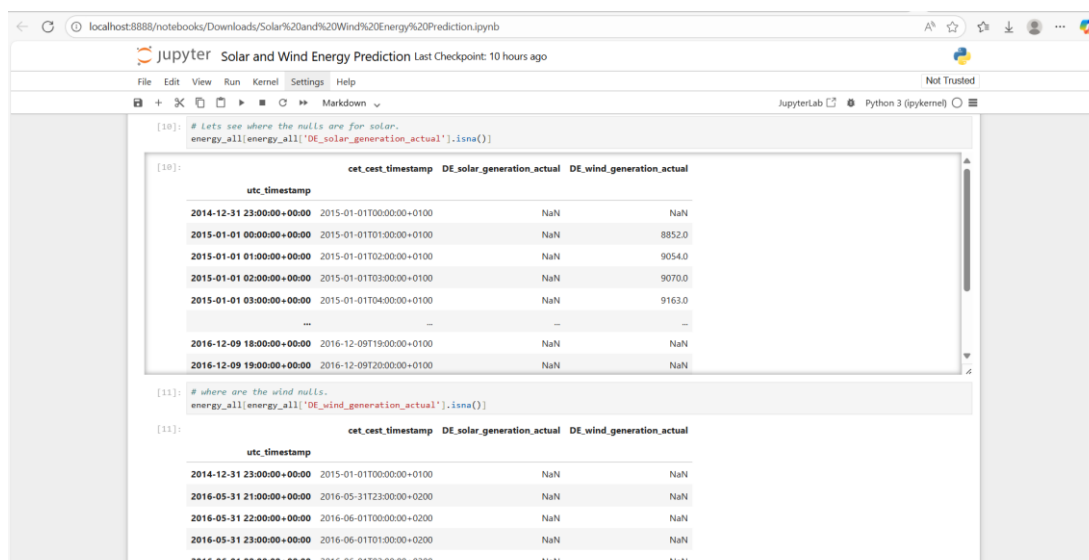


Рис. 3. Значення NaN серед часових рядів для сонячної та вітрової енергії

Джерело: побудовано авторами в Jupyter Notebook.

Це ускладнює перетворення даних під час моделювання. Наявність таких пропусків потребує етапу попередньої обробки даних, зокрема заповнення або коригування значень, аби забезпечити коректний статистичний аналіз і моделювання часових закономірностей. Не можна взяти логарифм нуля. Нулі виникають тому, що вночі та рано вранці відсутнє сонце.

Нульових значень не так і вже багато. Для коректного наступного моделювання потрібно заповнити їх значенням за день до цього. Це не дуже значний відсоток даних, і надалі усередним ці дані за весь день. Такий підхід виглядає більш прийнятним, оскільки це не спотворює прогнози.

Залишилися лише нульові значення з першого дня, оскільки решта була заповнена нульовими значеннями з попереднього дня. Це перетворило фіктивний стовпець `set_cest_timestamp` на нульові значення в тому ж місці. Саме тому він був використаний як фіктивний стовпець. Дані мають великий міжквартильний діапазон. Висновок узгоджується з фізичною природою процесу, оскільки сонячна радіація досягає максимуму близько полудня, що й призводить до великого міжквартильного діапазону в цей період (рис. 4).

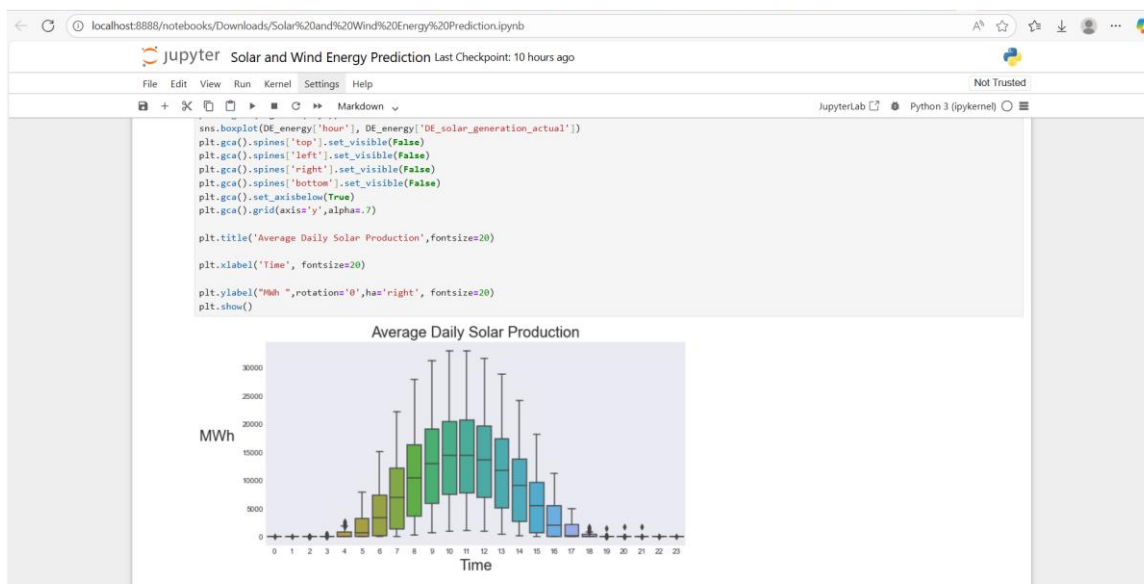


Рис. 4. Вох-plot усередненої кількості виробленої сонячної електроенергії  
Джерело: побудовано авторами в Jupyter Notebook.

Розподіл даних є асиметричним і змінюється протягом доби, тому цілком виправданим є використання робастних статистик, таких як медіана та IQR.

Прогнозування генерації сонячної енергії є ключовим завданням для балансування енергосистеми та планування навантажень. Зазвичай воно базується на історичних часових рядах, метеорологічних змінних (сонячна радіація, хмарність, температура) та добовій і сезонній циклічності. Для розв'язання задачі застосовують статистичні підходи (ARIMA, SARIMA), методи машинного навчання (Random Forest, XGBoost) і глибокі нейронні мережі (LSTM, GRU).

У процесі прогнозування важливо коректно обробляти нульові значення та враховувати нічні години, коли генерація фізично відсутня.

Зазвичай дані впорядковують у часовому порядку та розділяють на навчальну, валідаційну й тестову частини без випадкового перемішування.

До навчальної вибірки можуть входити ознаки години доби, дня тижня, пори року, а також метеорологічні параметри, зокрема сонячна радіація, хмарність, температура повітря та вологість, тоді як цільовою змінною є обсяг згенерованої сонячної енергії.

Інтеграція супутникових та метеорологічних даних у прогнозуванні генерації сонячної енергії відбувається через поєднання різнорідних джерел інформації для побудови більш точних моделей.

Особливу увагу варто приділити коректній інтерпретації нульових значень, щоб навчальна вибірка не вводила модель в оману. Для підвищення якості моделей у навчальну вибірку включають часові ознаки, такі як година доби, день тижня та номер місяця. Також ефективним підходом є використання лагових значень генерації, що дозволяє моделі враховувати попередній стан системи. Навчальна вибірка для даних по сонячній енергії і характер розподілу що побудовані авторами в Jupyter Notebook представлені відповідно на рис. 6 та рис. 7.

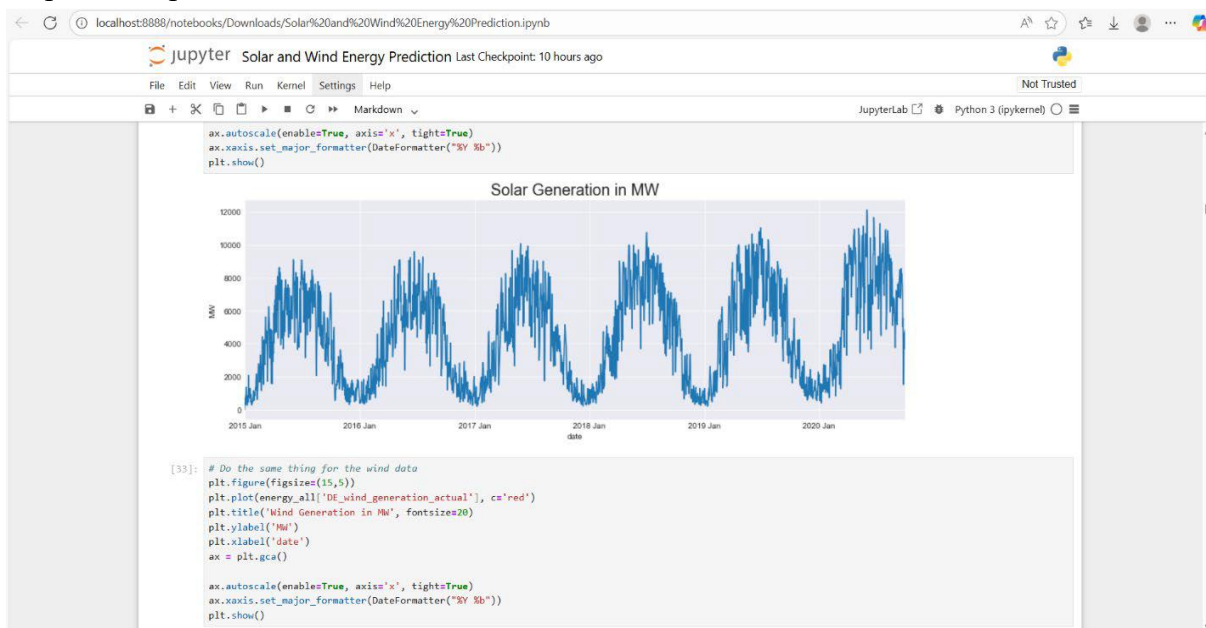


Рис. 6. Навчальна вибірка для даних по сонячній енергії

Джерело: побудовано авторами в Jupyter Notebook.

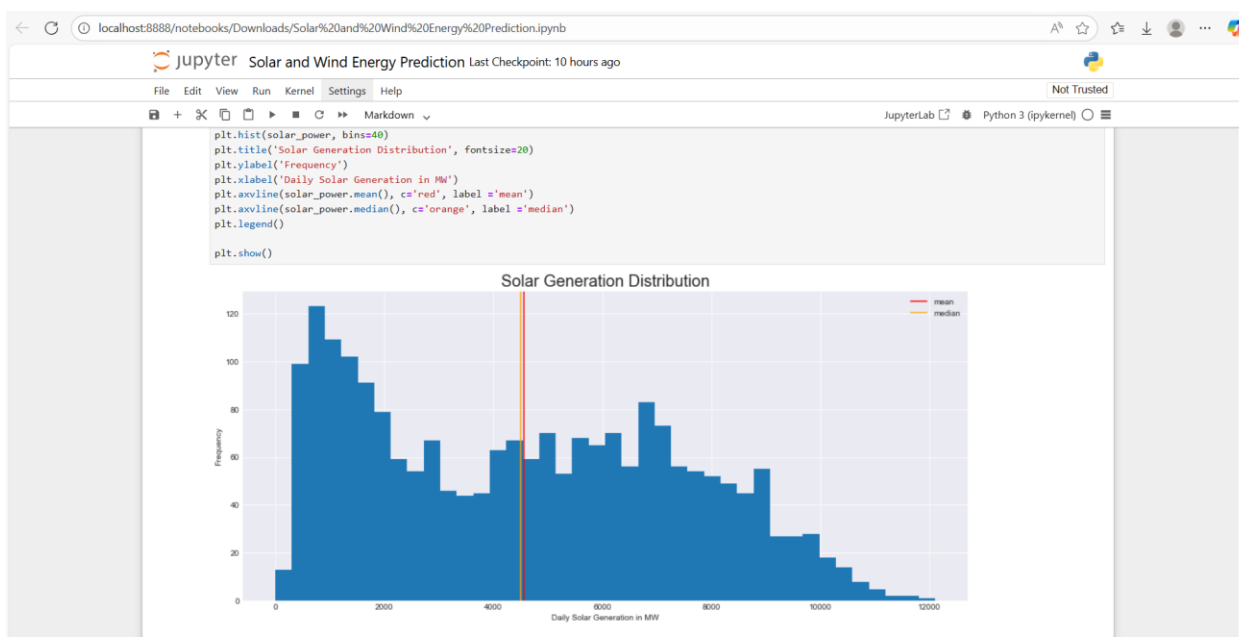


Рис. 7. Характер розподілу сонячної енергії

Джерело: побудована авторами в Jupyter Notebook.

Навчальну вибірку слід відокремлювати від тестової та валідаційної частин з урахуванням часової послідовності. Такий підхід запобігає витоку інформації з майбутнього та забезпечує коректну оцінку якості прогнозування. Якісно сформована навчальна вибірка є запорукою точного прогнозування генерації сонячної енергії. Урахування фізичних закономірностей, сезонності та особливостей даних дозволяє значно підвищити надійність моделей машинного навчання.

Для візуалізації розподілу основних статистичних характеристик, таких як медіана, квартилі та можливі викиди, скористаємося box-plot (рис. 8), який допомагає швидко оцінити, як розподілені дані генерації сонячної енергії, а також виявити асиметрію або наявність аномальних значень.

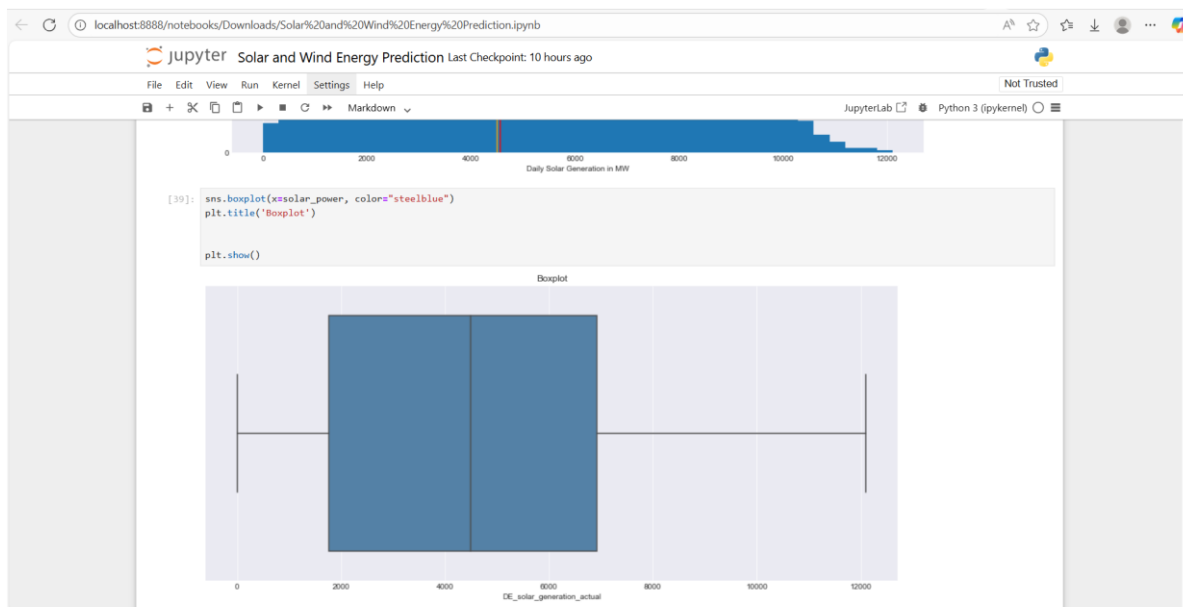


Рис. 8. Box-plot розподілу частоти різних об'ємів генерації сонячної енергії  
Джерело: побудовано авторами в Jupyter Notebook.

Box-plot (рис. 8 і рис. 9) представляє інтерквартильний діапазон (IQR), який охоплює 50% найсередніших значень. Верхня частина ящика – це третій квартиль (Q3), а нижня – перший квартиль (Q1).

IQR – число, яке показує розкид середньої половини (тобто середні 50%) набору даних і допомагає визначити викиди. IQR – це різниця між Q3 і Q1.

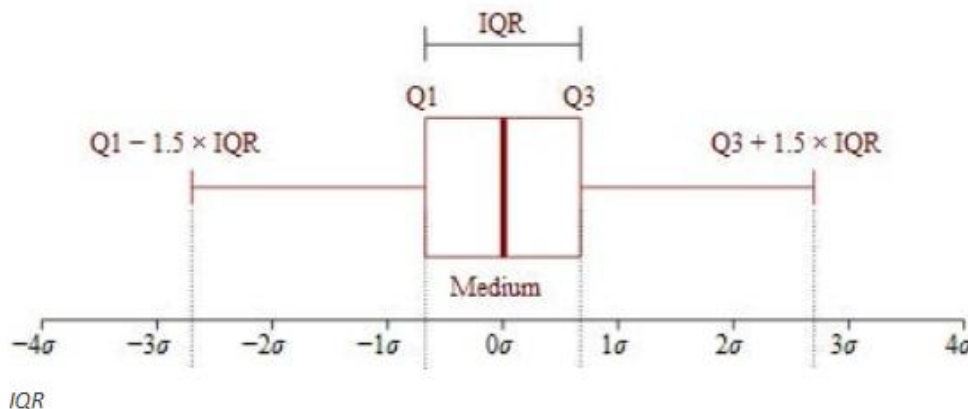


Рис. 9. Інтерквартильний розмах

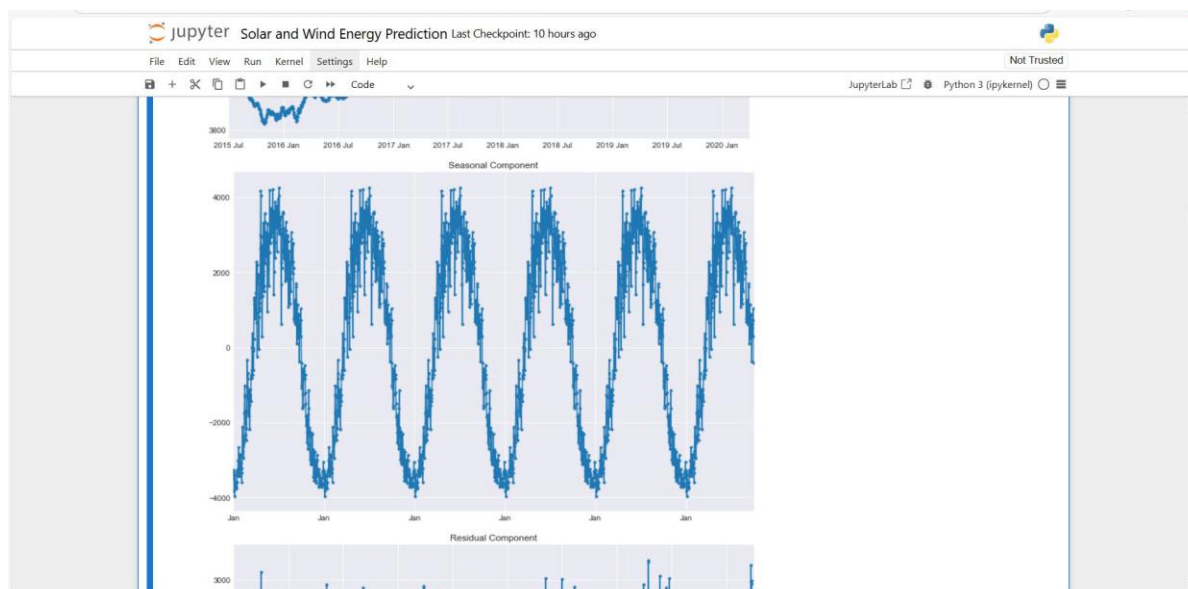
Лінія всередині ящика: показує медіану (Q2) даних. Вуси: виходять з ящика і зазвичай простягаються до найменшого та найбільшого значення, які не є викидами. Викиди часто показуються як окремі точки. Викиди: значення, що виходять за межі 1,5 раза IQR від країв ящика.

Для моделювання був використаний підхід до статистичного прогнозування часових рядів. Побудовано графіки автокореляції та часткової автокореляції. На рисунках нижче наведено вилучення тренду (рис. 10), сезонної (рис. 11) і шумової компоненти (рис. 12).



*Рис. 10. Тренд генерації сонячної енергії*

Джерело: побудований авторами в Jupyter Notebook.



*Рис. 11. Сезонні компоненти*

Джерело: побудований авторами в Jupyter Notebook.

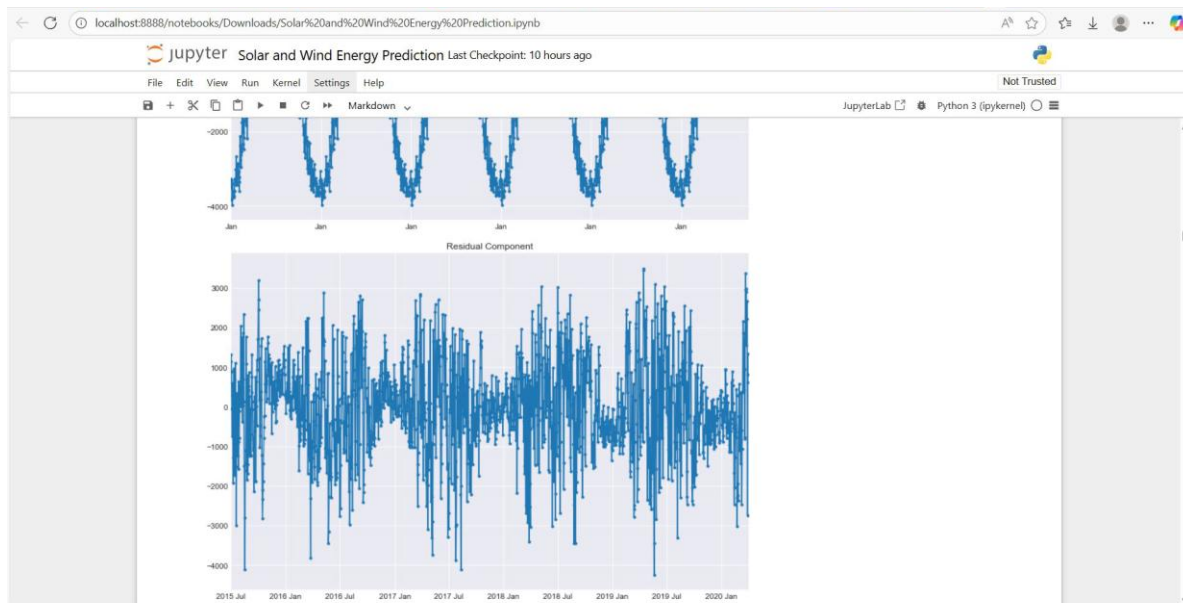


Рис. 11. Компонента білого шуму

Джерело: побудований авторами в Jupyter Notebook.

Наведені три графіки автокореляції свідчать що після декомпозиції часового ряду (другий та третій графіки автокореляція при мінімальному кроці стає практично нульовою (рис. 12).

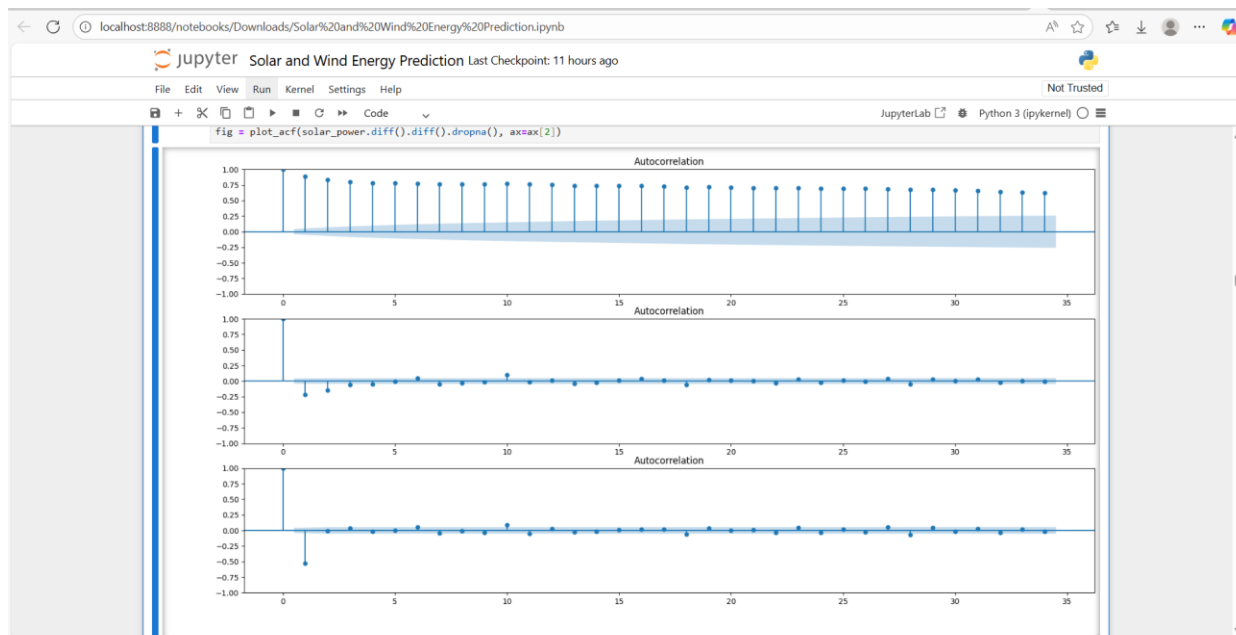


Рис. 12. Результати автокореляції для повного сигналу, сезонної і шумової компонент

Джерело: побудований авторами в Jupyter Notebook.

Далі був використаний модуль python scikit learnt для створення рекурентної моделі, яка ідеально підходить для розв’язання такої задачі і здійснили її навчання. Модель оптимізувалась досить швидко. На рис. 13 наведено результати прогнозування на тестовій вибірці за довгий період.

Останнім кроком будуємо довірчі границі (рис. 14), які свідчать що 80 % даних лежить у межах довірчих границь. Проте дані мають сезону гетероскедастчність (дисперсія

(варіація) даних змінюється залежно від часу або сезону), а саме в літні дні спостерігається набагато більше коливання генерації сонячної енергії, ніж взимку.

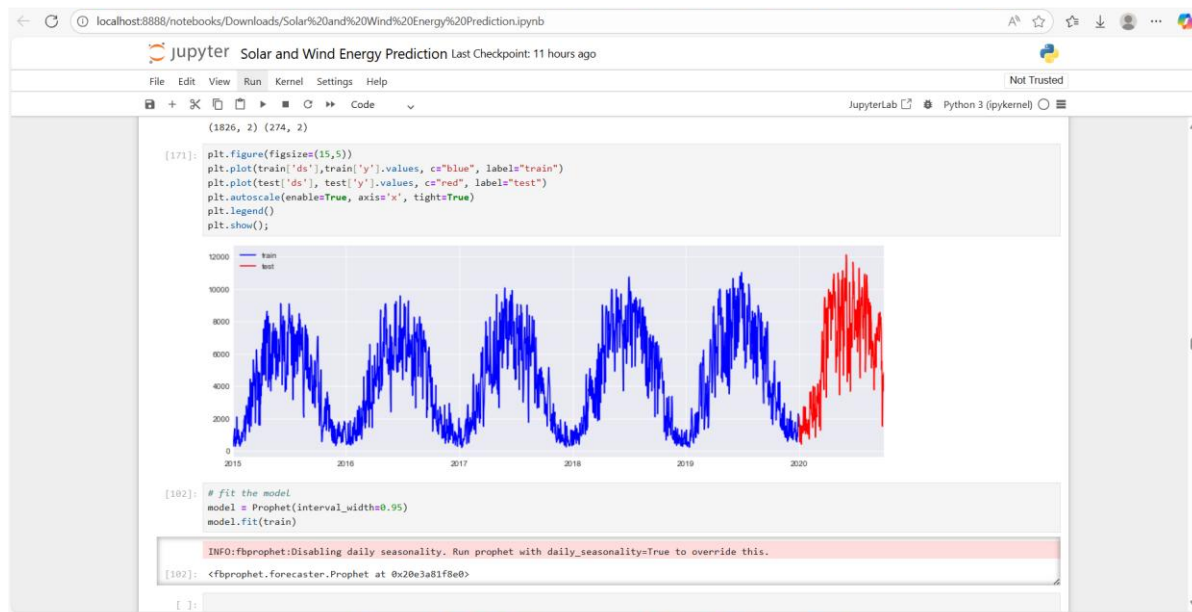


Рис. 13. Побудова прогнозу на один рік для генерації сонячної електроенергії Побудований авторами в Jupyter Notebook.

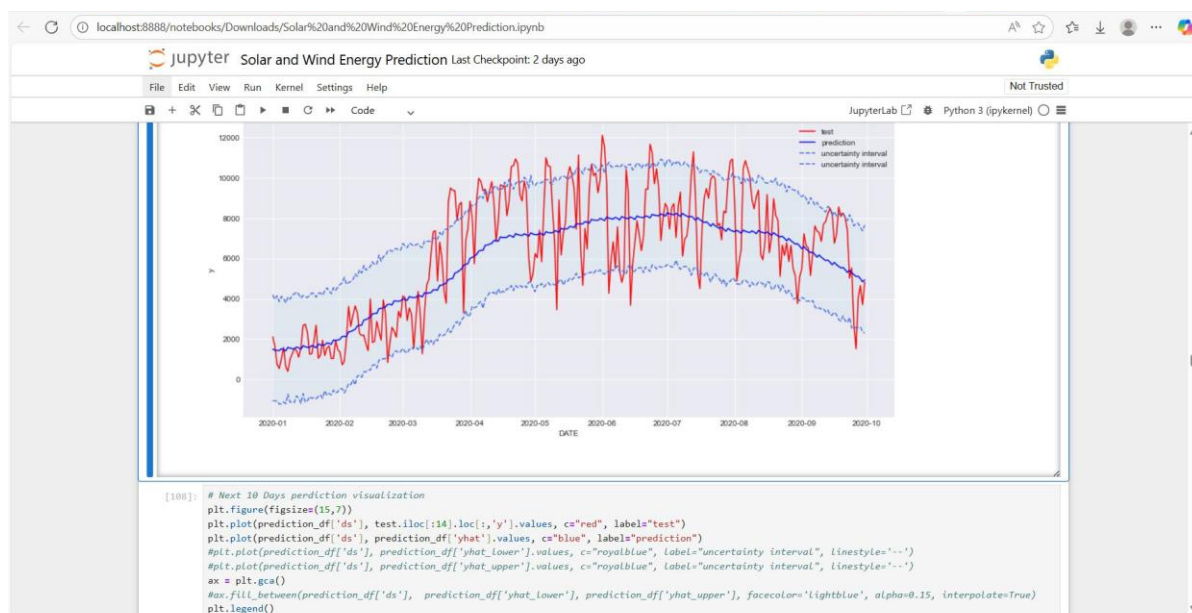


Рис. 14. Побудова довірчих границь

**Висновки.** Використання технологій Big Data в альтернативній енергетиці відкриває нові можливості для підвищення ефективності, прогнозованості та надійності енергосистем. Обробка великих масивів даних дозволяє створювати точні моделі поведінки енергогенерації та оптимізувати балансування навантажень. Отримані результати демонструють практичне впровадження цих принципів і підтверджує потенціал аналітики великих даних у сфері відновлюваної енергетики.

Для моделювання використано підхід до статистичного прогнозування часових рядів. Побудовано графіки автокореляції та часткової автокореляції. Моделі оцінювались за метриками MAE та RMSE. В роботі розглянуто та проаналізовано набір даних OPSD

Germany, здійснено його комплексний аналітичний аналіз. Побудовано моделі класу RNN та здійснено їх навчання за допомогою модуля Python scikit-learn. Результати показали, що застосування RNN забезпечує високу точність 86,191 % для передбачення генерації сонячної енергії в короткостроковому горизонті, враховуючи сезонні та трендові компоненти. Зростання генерації сонячної й вітрової енергії чітко корелює з розширенням інфраструктури відновних джерел енергетики у Німеччині. Варто звернути увагу що сонячна електрогенерація має підвищену дисперсію даних в літні періоди і навпаки дисперсія знижується взимку.

Отримане прогнозування є прикладом практичного застосування технологій Big Data та аналітики часових рядів у системах управління відновлюваною енергетикою. Вона демонструє, як інтеграція відкритих енергетичних даних, хмарних обчислень і машинного навчання може підвищити точність прогнозів, необхідних для планування навантажень, стабільності мережі та розробки політики сталого розвитку.

### Список використаних джерел

1. Li, H., Wang, L., Liu, J., & Zhou, A. (2020). Short-term solar power forecasting using deep learning models. *Renewable Energy*, 152, 635–645.
2. Ahmad, F., Khan, M. A., & Lee, S. (2020). Machine learning techniques for solar irradiance prediction: A review. *Solar Energy*, 209, 224–245.
3. Zhang, Y., Li, Z., & Chen, X. (2018). Satellite data and meteorological parameters integration for PV power forecasting. *Applied Energy*, 230, 1056–1067.
4. Jain, P., & Kumar, R. (2020). Ensemble learning methods for solar power forecasting: A comparative study. *Energy Reports*, 6, 1205–1218.
5. Rahman, M., Chowdhury, S., & Rahman, T. (2020). IoT-based real-time monitoring and predictive maintenance of solar panels. *Renewable and Sustainable Energy Reviews*, 132, 1100–1113.
6. Park, J., & Kim, K. (2020). Big Data analytics for forecasting solar energy generation in smart grids. *IEEE Access*, 8, 104563–104574.
7. Singh, A., & Gupta, N. (2021). Challenges and opportunities in machine learning-based solar energy prediction. *Energy Reports*, 7, 1120–1135.
8. Письменна У. та ін. (2021). Становлення ринкового механізму стимулювання розвитку систем накопичення енергії. *Вісник СумДУ. Серія «Економіка»*, (3), 31–39.
9. Станицина В. І. та ін. (2023). Можливості підвищення енергетичної незалежності територіальних громад в Україні. *Енергетика та громади* (с. 12–20).
10. Кизим М. І. та ін. (2023). Енергоефективність та енергозалежність в Україні з урахуванням світових тенденцій. *Проблеми економіки*, (2(56)), 41–55.
11. Кудря, С. О. (ред.). (2025). *Відновлювані джерела енергії: видання третє, оновлене*. Інститут відновлюваної енергетики НАНУ.
12. Chandel S.S. et al. Review of deep learning techniques for power generation prediction of industrial solar photovoltaic plants // *Solar Compass* 8 (2023) <http://www.elsevier.com/locate/solcom>
13. Iheanetu Kelachukwu J. Solar Photovoltaic Power Forecasting: A Review <https://www.mdpi.com/journal/sustainability>.
14. Bedi, J., & Toshniwal, D. (2022). Deep learning framework to forecast electricity demand. *Energy*, 239, 122006.
15. Ahmad, T., Zhang, D., & Dai, H. (2020). Toward smart energy systems — Big data and machine learning approaches. *Renewable and Sustainable Energy Reviews*, 122, 109772.
16. Cavazzi, S., & Dalla Longa, F. (2019). GIS-based assessment of onshore wind energy potential in Europe. *Renewable Energy*, 139, 734–747.
17. Khan, A., et al. (2020). Big data analytics and smart grid integration: A comprehensive review. *IEEE Access*, 8, 198947–198967.
18. Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56, 215–225.
19. Siano, P., De Marco, G., Rolán, A., & Loia, V. (2019). A survey and evaluation of the potentials of artificial intelligence applications in power systems. *IEEE Transactions on Power Systems*, 34(5), 3836–3846.

20. Open Power System Data. 2020. *Data Package Time series*. Version 2020-10-06. [https://doi.org/10.25832/time\\_series/2020-10-06](https://doi.org/10.25832/time_series/2020-10-06).

### References

1. Li, H., Wang, L., Liu, J., & Zhou, A. (2020). Short-term solar power forecasting using deep learning models. *Renewable Energy*, 152, 635–645.
2. Ahmad, F., Khan, M. A., & Lee, S. (2020). Machine learning techniques for solar irradiance prediction: A review. *Solar Energy*, 209, 224–245.
3. Zhang, Y., Li, Z., & Chen, X. (2018). Satellite data and meteorological parameters integration for PV power forecasting. *Applied Energy*, 230, 1056–1067.
4. Jain, P., & Kumar, R. (2020). Ensemble learning methods for solar power forecasting: A comparative study. *Energy Reports*, 6, 1205–1218.
5. Rahman, M., Chowdhury, S., & Rahman, T. (2020). IoT-based real-time monitoring and predictive maintenance of solar panels. *Renewable and Sustainable Energy Reviews*, 132, 1100–1113.
6. Park, J., & Kim, K. (2020). Big Data analytics for forecasting solar energy generation in smart grids. *IEEE Access*, 8, 104563–104574.
7. Singh, A., & Gupta, N. (2021). Challenges and opportunities in machine learning-based solar energy prediction. *Energy Reports*, 7, 1120–1135.
8. Pysmenna, U., et al. (2021). Stanovlennia rynkovoho mekhanizmu stymuliuвання rozvytku system nakopychennia enerhii [Formation of the market mechanism for stimulating the development of energy storage systems]. *Visnyk SumDU. Serii «Ekonomika» – Visnyk of Sumy State University. Series: Economics*, (3), 31–39.
9. Stanytsyna, V. I., et al. (2023). Mozhyvosti pidvyshchennia enerhetychnoi nezalezhnosti terytorialnykh hromad v Ukraini [Opportunities to increase the energy independence of territorial communities in Ukraine]. *Enerhetyka ta hromady – Energy and Communities*, 12–20.
10. Kyzym, M. I., et al. (2023). Enerhoefektyvnist ta enerhozalezhnist v Ukraini z urakhu-vanniam svitovykh tendentsii [Energy efficiency and energy dependence in Ukraine considering global trends]. *Problemy ekonomiky – Problems of Economy*, 2(56), 41–55.
11. Kudria, S. O. (Ed.). (2025). *Vidnovliuvani dzherela enerhii: vydannia tretie, onovlene [Renewable energy sources]* (3rd updated ed.). Institute of Renewable Energy of the NAS of Ukraine.
12. Chandel, S. S., et al. (2023). Review of deep learning techniques for power generation prediction of industrial solar photovoltaic plants. *Solar Compass*, 8. <http://www.elsevier.com/locate/solcom>
13. Iheanetu, K. J. Solar photovoltaic power forecasting: A review. *Sustainability*. <https://www.mdpi.com/journal/sustainability>
14. Bedi, J., & Toshniwal, D. (2022). Deep learning framework to forecast electricity demand. *Energy*, 239, 122006.
15. Ahmad, T., Zhang, D., & Dai, H. (2020). Toward smart energy systems—Big data and machine learning approaches. *Renewable and Sustainable Energy Reviews*, 122, 109772.
16. Cavazzi, S., & Dalla Longa, F. (2019). GIS-based assessment of onshore wind energy potential in Europe. *Renewable Energy*, 139, 734–747.
17. Khan, A., et al. (2020). Big data analytics and smart grid integration: A comprehensive review. *IEEE Access*, 8, 198947–198967.
18. Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56, 215–225.
19. Siano, P., De Marco, G., Rolán, A., & Loia, V. (2019). A survey and evaluation of the potentials of artificial intelligence applications in power systems. *IEEE Transactions on Power Systems*, 34(5), 3836–3846.
20. Open Power System Data. (2020). *Data package time series (Version 2020-10-06)*. [https://doi.org/10.25832/time\\_series/2020-10-06](https://doi.org/10.25832/time_series/2020-10-06)

Дата першого надходження статті до видання: 26.02.2026  
Дата прийняття статті до друку після рецензування: 14.03.2026

**Oleksandr Vasyliiev<sup>1</sup>, Serhii Zubenko<sup>2</sup>**

<sup>1</sup>PhD student, 2nd year of study, specialty 193 – Geodesy and Land Management  
Taras Shevchenko National University of Kyiv, Educational and Scientific Institute “Institute of Geology” (Kyiv, Ukraine)  
E-mail: [vasilyeva455@gmail.com](mailto:vasilyeva455@gmail.com). ORCID: <https://orcid.org/0009-0009-2177-7859>

<sup>2</sup>Master’s student, 2nd year of study, specialty 103 – Earth Sciences  
Taras Shevchenko National University of Kyiv, Educational and Scientific Institute “Institute of Geology” (Kyiv, Ukraine)  
E-mail: [szubnko02@gmail.com](mailto:szubnko02@gmail.com). ORCID: <https://orcid.org/0009-0008-1166-4693>

**BIG DATA TECHNOLOGIES IN FORECASTING SOLAR ENERGY GENERATION**

*The growing share of renewable energy sources (RES), particularly solar power, in the global energy balance necessitates highly accurate forecasting of electricity generation. The variable nature of solar radiation, along with seasonal and diurnal fluctuations, poses significant challenges to the stability and reliability of power systems. Big Data technologies enable the real-time processing of massive datasets, improving forecast accuracy and optimizing power system operations.*

*This paper examines modern Big Data technologies and their application in forecasting solar energy generation. It focuses on methods for collecting, processing, and analyzing large volumes of data from various sources, including meteorological parameters, solar panel performance metrics, and satellite observations. The study explores machine learning and artificial intelligence algorithms for building forecasting models, which improve the precision of electricity generation predictions and optimize the integration of solar energy into power grids.*

*Big Data facilitates the accurate forecasting of energy production from solar power plants, accounting for weather conditions and seasonal variability. Furthermore, it enhances the efficiency of power grid management by enabling real-time balancing of supply and demand.*

*The proposed approaches demonstrate the effectiveness of using Big Data analytics to support decision-making in the renewable energy sector and ensure power grid stability. The findings of this study are applicable to production planning and resource management at solar power plants.*

*The temporal patterns of electricity production from RES are analyzed based on solar energy generation data. The primary focus is on the diurnal, seasonal, and interannual variability of generation, as well as the impact of environmental and climatic factors on production stability.*

**Keywords:** Big Data; solar energy production optimization; forecasting.

Table: 1. Fig.: 14. References: 20.